DOCUMENT RESUME

ED 447 959 PS 029 068

AUTHOR	Ponza, Michael; Briefel, Ronette; Corson, Walter; Devaney, Barbara; Glazerman, Steven; Gleason, Philip; Heaviside, Sheila; Kung, Susanna; Meckstroth, Alicia; Murphy, J. Michael; Ohls, Jim
TITLE	Universal-Free School Breakfast Program Evaluation Design Project: Final Evaluation Design.
INSTITUTION	Mathematica Policy Research, Princeton, NJ.
SPONS AGENCY	Food and Nutrition Service (USDA), Alexandria, VA. Office of Analysis and Evaluation.
PUB DATE	1999-12-20
NOTE	266p.
CONTRACT	53-3198-7-006
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC11 Plus Postage.
DESCRIPTORS	*Breakfast Programs; *Children; Elementary Education; Elementary School Students; Evaluation Methods; Evaluation Problems; *Nutrition; Pilot Projects; Program Effectiveness; *Program Evaluation; *Research Design
IDENTIFIERS	*Universal School Breakfast Program; William F Goodling Child Nutrition Reauth Act 1998

ABSTRACT

The Child Nutrition Act of 1998 authorized demonstration pilot projects in up to six school food authorities and a rigorous evaluation to assess the effects of providing free school breakfasts to elementary school children. This report describes the evaluation strategy and data collection plans. Part 1 of the report provides background information, describes the school breakfast program (SBP), summarizes the literature on breakfast and learning, and discusses the rationale for the Universal-Free School Breakfast Program (USBP). This part also identifies the requirements of the demonstration and evaluation, provides a conceptual framework of how participation in a USBP may affect student outcomes, and provides an overview of the full set of evaluation design components. Part 2 describes the design for the impact study. Research questions and hypotheses to be tested are discussed, as are outcomes and measures that will be used to assess program effectiveness. This part also describes the approach for evaluating the USBP demonstration programs, discusses evaluation and sampling design issues and strategies, presents an analysis plan for assessing program impacts, and describes data collection plans. Part 3 presents the approach for conducting the implementation study of the USBP pilot project, describing study topics and research questions and outlining plans for process data collection. Part 4 provides an assessment of the feasibility of implementing the preferred evaluation design. This part reviews the statistical precision achieved, provides an estimate of the demonstration costs, and identifies strengths, weaknesses, and risks. (Contains 67 references.) (KB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PS

ED 447 959

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Contract No.: 53-3198-7-006
Order No.: 43-3198-8-3330
MPR Reference No.: 8640-600

Universal-Free School Breakfast Program Evaluation Design Project

Final Evaluation Design

December 20, 1999

Michael Ponza
Ronette Briefel
Walter Corson
Barbara Devaney
Steven Glazerman
Philip Gleason
Sheila Heaviside
Susanna Kung
Alicia Meckstroth
J. Michael Murphy
Jim Ohls

PS 029068

Submitted to:

USDA, Food and Nutrition Service
Office of Analysis, Nutrition, and Evaluation
3101 Park Center Drive, 5th Floor
Alexandria, VA 22302

Project Officer:
John Endahl

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535
Project Director:
Michael Ponza
Principal Investigators:
Barbara Devaney
J. Michael Murphy

BEST COPY AVAILABLE

CONTENTS

Chapter		Page
	PART 1: INTRODUCTION AND BACKGROUND	
I	INTRODUCTION	1
	A. THE USBP DEMONSTRATION AND EVALUATION	2
	B. CHALLENGES TO THE USBP EVALUATION	3
	1. The "Dilution" Effect	4
	2. Variation Between Schools	5
	C. OVERVIEW OF THE REPORT	6
II	BACKGROUND	7
	A. OVERVIEW OF THE SCHOOL BREAKFAST PROGRAM	7
	B. EFFECTS OF BREAKFAST CONSUMPTION ON CHILDREN'S LEARNING: A SUMMARY OF THE LITERATURE	9
	1. Nutrition and Cognitive Development	10
	2. Eating Breakfast, Daily Nutrient Intake, and Behavioral and Cognitive Development	12
	3. School Breakfasts, Dietary Status, and School Performance and Achievement	13
	C. RATIONALE FOR THE USBP PILOT PROJECTS	18
	D. DEMONSTRATION AND EVALUATION REQUIREMENTS	20
	1. Discussion	22
	2. Selecting SFAs	23
	3. Required SFA Application Data	24
	E. CONCEPTUAL FRAMEWORK	25
	F. OVERVIEW OF THE USBP EVALUATION	29
	1. Implementation Analysis	30
	2. Impact Analysis	30
	G. SUMMARY	33

CONTENTS *(continued)*

Chapter		Page
	PART 2: THE IMPACT EVALUATION DESIGN	
III	IMPACT STUDY RESEARCH QUESTIONS AND HYPOTHESES	35
	A. IMPACT STUDY RESEARCH OBJECTIVES AND QUESTIONS	35
	B. PRINCIPAL HYPOTHESES THAT THE IMPACT ANALYSIS WILL ADDRESS	37
	1. School Breakfast Program Participation	40
	2. Breakfast Consumption Patterns	40
	3. Dietary Intake	42
	4. Food Security	43
	5. Attendance	43
	6. Cognitive Functioning and Behavior	44
	7. Children's Health Status	45
	8. School Climate	45
	9. Academic Achievement	46
IV	DATA DOMAINS AND OUTCOME MEASURES	49
	A. SCHOOL BREAKFAST PARTICIPATION OUTCOMES	50
	1. School Breakfast Definition	51
	2. Target Day Participation and Usual Participation	51
	3. Food Selected and Food Consumed	53
	B. DIETARY OUTCOMES	53
	1. Breakfast Consumption	54
	2. Dietary Intake	54
	3. Food Security	58
	C. INTERMEDIATE SCHOOL OUTCOMES	59
	1. Attendance and Tardiness	60
	2. Classroom Behavior and Disciplinary Incidents	61
	3. Cognitive Functioning	64
	4. Emotional Effects	67

CONTENTS *(continued)*

Chapter		Page
IV (continued)	D. LONG-TERM SCHOOL ACHIEVEMENT OUTCOMES	68
	1. Achievement Domains	70
	2. Achievement Test Instruments	71
	3. Who Should Administer the Test?	75
	4. When to Measure Achievement	77
	5. Grade Levels	78
	E. HEALTH OUTCOMES	79
	1. Self-Reported Health	80
	2. Number of Visits to the School Nurse	80
	3. Absences Due to Illness	80
	4. Height and Weight	81
	F. SCHOOL-WIDE OUTCOMES	81
	1. School Climate	81
	2. Attitudes	82
	G. PROPOSED CONTROL VARIABLES	82
	1. Key Determinants/Predictors of Individual Cognitive Achievement	83
	2. Key Determinants of Attendance and Tardiness	84
	3. Key Determinants of Individual Nutrient Intake and Health Outcomes	85
	4. School-Level Variables	86
V	RANDOM ASSIGNMENT DESIGN	87
	A. THE CASE FOR AN EXPERIMENTAL DESIGN	89
	1. Random Assignment Designs	90
	2. Nonexperimental Design Alternatives	91
	3. The Importance of Obtaining Reliable Evaluation Results Favors Using a Random Assignment Design to Evaluate the USBP Pilot Projects	95

CONTENTS *(continued)*

Chapter		Page
V (continued)	B. FEATURES OF THE RANDOM ASSIGNMENT DESIGN	96
	1. Preferred Approach Is to Randomize Schools	96
	2. The Design Does Not Include Planned Variation in the Demonstration Treatment	100
	3. The Regular SBP as the Counterfactual	102
	4. The Evaluation Will Collect School-Level and Student-Level Data	103
	5. The Evaluation Will Exclude Kindergarten Students from the Survey Research Sample	104
	6. The Evaluation Will Include a Longitudinal Component	106
	7. Length of the Follow-up Period	112
	C. STATISTICAL PRECISION AND POWER	115
	1. Design Options	115
	2. Assumptions	117
	3. Factors Affecting Precision Levels	119
	4. Detailed Precision Estimates of the Effects of the USBP on Student Outcomes	120
VI	DATA COLLECTION PLANS	137
	A. OVERVIEW OF DATA COLLECTION INSTRUMENTS/ METHODS	137
	B. SCHOOL RECORDS DATA COLLECTION	140
	1. Methods	142
	2. Timing of Data Collection	143
	3. Issues	145
	C. DATA COLLECTION FROM STUDENTS AND PARENTS	147
	1. Dietary Intake	147
	2. Student/Parent Surveys	156
	3. Achievement Tests	159
	4. Same-Day Cognitive Tests	163

CONTENTS *(continued)*

Chapter		Page
VI (continued)	D. DATA COLLECTION FROM TEACHERS (ON THE SAMPLED STUDENTS)	165
	1. Methods	166
	2. Timing of Data Collection	166
	3. Issues	167
	E. OBSERVATION OF SCHOOL MEALS	167
	1. Method	167
VII	IMPACT ANALYSIS	169
	A. TYPES OF ANALYSIS	169
	1. Descriptive Analysis	170
	2. Estimation of the Effects of USBP Availability on Student Outcomes	171
	3. Estimation of the Effects of Breakfast Program Participation on Student Outcomes	173
	B. UNIT OF ANALYSIS	174
	1. School-Level Analysis	175
	2. Student-Level Analysis	176
	C. MODELS TO BE ESTIMATED	176
	1. Model I: School-Level Model of the Effects of USBP Availability	177
	2. Model II: Student-Level Model of the Effects of USBP Availability	178
	3. Model III: School-Level Model of the Effects of Breakfast Program Participation	181
	4. Model IV: Student-Level Model of the Effects of Breakfast Program Participation	183

CONTENTS *(continued)*

Chapter		Page
VII (continued)	D. SELECTION BIAS	185
	1. The Problem	186
	2. Controlling Directly for Relevant Factors	188
	3. Fixed-Effects Models	189
	4. Instrumental Variables/Selection Correction Models	192
	5. Controlling for Usual Participation	194
	PART 3: IMPLEMENTATION STUDY DESIGN	
VIII	IMPLEMENTATION STUDY DESIGN	197
	A. RESEARCH ISSUES	198
	B. RESEARCH TOPICS	199
	1. Organizational Aspects	203
	2. Service Delivery	204
	3. Meals Offered	205
	4. Program Costs and Administrative Burden	205
	5. Participation	207
	6. School Environment	208
	C. DATA COLLECTION PLANS	208
	1. Semistructured Telephone Interviews with SFA Administrators	216
	2. Semistructured Telephone Interviews with Cafeteria Managers	216
	3. Semistructured Telephone Interviews with School Administrators and Other Key Staff	217
	4. Meals Offered	218
	5. Program Costs	220
	6. Focus Groups	222
	7. Program Documents and Records	223
	D. DATA ANALYSIS AND SYNTHESIS	224
	1. Analysis of Quantitative Data	225
	2. Analysis of Qualitative Data	225
	3. Integration of Quantitative and Qualitative Data	226

CONTENTS *(continued)*

Chapter		Page
	PART 4: FEASIBILITY OF IMPLEMENTATION	
IX	OVERALL FEASIBILITY OF EVALUATION DESIGN OPTIONS	227
	A. OVERVIEW OF THE DESIGN OPTIONS	227
	1. Design 1a: Randomize 144 Schools, No Preimplementation Survey	230
	2. Design 1b: Randomize 144 Schools, Conduct Preimplementation Survey	231
	3. Design 2a: Randomize 120 Schools, No Preimplementation Survey	232
	4. Design 2b: Randomize 120 Schools, Conduct Preimplementation Survey	232
	B. PRECISION OF THE IMPACT ESTIMATES	233
	1. Participation in the USBP	233
	2. Dietary Intake at Breakfast	233
	3. Student Academic Achievement	234
	C. DEMONSTRATION AND EVALUATION COSTS	234
	D. POTENTIAL FOR SUCCESSFUL COMPLETION	235
	1. Risks of Implementing Design 1a and Design 2a	235
	2. Risks of Implementing Design 1b and Design 2b	235
	E. RECOMMENDED APPROACH	236
	REFERENCES	239

TABLES

Table		Page
III.1	OVERVIEW OF RESEARCH QUESTIONS, OUTCOME MEASURES, AND UNIT OF MEASUREMENT FOR THE USBP EVALUATION'S IMPACT STUDY	38
III.2	EXPECTED EFFECTS OF THE USBP ON STUDENT OUTCOMES	42
IV.1	INSTRUMENTS FOR MEASURING SHORT-TERM BEHAVIORAL OUTCOMES	62
IV.2	INSTRUMENTS FOR MEASURING SHORT-TERM COGNITIVE AND EMOTIONAL OUTCOMES	67
V.1	OVERVIEW OF THE STATISTICAL PRECISION OF THE ALTERNATIVE IMPACT EVALUATION DESIGNS	121
V.2	MINIMUM DETECTABLE IMPACTS FOR THE UNIVERSAL-FREE SCHOOL BREAKFAST PROGRAM EVALUATION: PARTICIPATION EFFECTS	123
V.3	MINIMUM DETECTABLE USBP IMPACTS ON DIETARY INTAKE WITHOUT PRE-IMPLEMENTATION SURVEY	126
V.4	MINIMUM DETECTABLE DIFFERENCES IN DIETARY INTAKE UNDER DESIGN 1B BASED ON COMPARISONS OF POTENTIAL PARTICIPANTS	129
V.5	MINIMUM DETECTABLE USBP IMPACTS ON STUDENT ACHIEVEMENT WITHOUT PREIMPLEMENTATION SURVEY	132
V.6	MINIMUM DETECTABLE DIFFERENCES IN STUDENT ACHIEVEMENT UNDER DESIGN 1B BASED ON COMPARISONS OF POTENTIAL PARTICIPANTS	133
VI.1	USBP DEMONSTRATION EVALUATION DATA COLLECTION INSTRUMENTS/METHODS AND DOMAINS FOR THE IMPACT STUDY	138
VI.2	USBP EVALUATION: ILLUSTRATIVE DATA COLLECTION PLANS FOR THE IMPACT STUDY	144

TABLES *(continued)*

Table		Page
VI.3	OVERVIEW OF DATA COLLECTION INSTRUMENTS FOR THE USBP IMPACT EVALUATION	148
VIII.1	IMPLEMENTATION STUDY TOPICS AND DATA SOURCES	200
VIII.2	IMPLEMENTATION STUDY: KEY RESEARCH TOPICS AND EXAMPLES OF QUESTIONS	209
IX.1	SUMMARY OF UNIVERSAL-FREE SCHOOL BREAKFAST PROGRAM IMPACT EVALUATION DESIGN OPTIONS	228

FIGURES

Figure		Page
II.1	CONCEPTUAL FRAMEWORK FOR ANALYZING THE EFFECTS OF A UNIVERSAL-FREE SCHOOL BREAKFAST PROGRAM	27
V.1	DATA COLLECTION DESIGN: STUDENT SAMPLES	109

PART 1:

INTRODUCTION AND BACKGROUND

I. INTRODUCTION

The School Breakfast Program (SBP), authorized by the Child Nutrition Act of 1966, started as a pilot program to provide funding for school breakfasts in poor areas and areas where children had to travel great distances to school. The intent was to provide a nutritious breakfast to children who might otherwise not receive one. The importance of a nutritious breakfast is supported by several studies that have linked breakfast to improved dietary status and enhanced school performance. More recent research suggests that providing school breakfasts to low-income children is associated with the greater likelihood of eating a substantive breakfast, improved school attendance, and decreased tardiness (Pollitt and Mathews 1998; and Briefel et al. 1999). Less is known about the effects of school breakfasts on children's cognitive functioning and academic achievement, although some studies suggest that school breakfasts may lead to improvements in these outcomes (Briefel et al. 1999).

In response to the body of evidence suggesting that school breakfasts provide dietary and educational benefits, many observers have urged that school breakfasts be more widely available. Within this context, Congress passed Section 109 of the William F. Goodling Child Nutrition Act of 1998 (P.L. 105-336). This authorizes the Secretary of Agriculture, through the Food and Nutrition Service (FNS), U.S. Department of Agriculture (USDA), to conduct a demonstration and evaluation that will rigorously assess the effects of the Universal-Free School Breakfast Program (USBP) on program participation and a broad range of student outcomes, including academic achievement, school attendance and tardiness, classroom behavior and attentiveness, and dietary status.

A. THE USBP DEMONSTRATION AND EVALUATION

The Child Nutrition Act of 1998 authorized demonstration pilot projects in up to six school food authorities (SFAs) and a rigorous evaluation to assess the effects of providing free school breakfasts to elementary school children. The evaluation is to address four main research objectives:

1. Determine the effect that participation by elementary schools in the USBP pilot projects has on student participation in the breakfast program, on the costs to the federal government in increases in funding for reimbursing program meals, and on simplifying paperwork the schools must complete and other administrative procedural participation requirements

2. Assess the effects of the USBP in elementary schools on selected student outcomes, including academic achievement, school attendance and tardiness, classroom behavior and discipline, and dietary intakes over the course of a day

3. Compare the characteristics and outcomes of USBP participants with those of nonparticipating children

4. Compare the characteristics and outcomes of USBP participants with children participating in the regular SBP

FNS has contracted with Mathematica Policy Research, Inc. to develop a comprehensive, rigorous study design for evaluating the USBP pilot programs. This report describes the evaluation strategy, which includes a methodologically rigorous design, a well-defined set of outcome variables, and data collection plans comprehensive enough to allow the gathering of evidence on program effectiveness.

The evaluation strategy is a randomized design: it first pairs schools, then randomly assigns each of the schools in each pair to either a treatment (USBP) or control (regular SBP) group. Next, samples of students are drawn from each school. The evaluation plan involves a wide range of data collection activities, including the collection of school administrative data, surveys of students and their parents, extraction of school records data, administration of cognitive tests and student

achievement tests, and site visits. Because the actual school districts applying for the demonstration are unknown at this time, the design considers two possible sample sizes for the number of schools. The first version assumes student achievement data will be available at two or more points in time for sampled students from tests routinely administered by the school districts chosen for the study. The second version would use district-administered achievement tests to supply baseline data and would obtain follow-up data on student achievement through tests administered by the evaluation contractor. In addition, because it is uncertain whether the demonstration will start in September 2000 or in January 2001, the design includes an option to conduct a preimplementation survey of students' parents. This survey would only be conducted if demonstration implementation is delayed until January 2001.

B. CHALLENGES TO THE USBP EVALUATION

Several challenges to the USBP evaluation must be addressed if the design is to yield valid and policy-relevant evidence on the effects of providing free school breakfasts to elementary school children. One challenge warrants special attention: whether, given the resources available for the demonstration and evaluation, the preferred design has the statistical precision to estimate impacts on student outcomes, especially student achievement. Because the USBP is generally considered an intervention at the school level, the preferred approach to evaluating the program is to use schools as the unit of assignment. Even under the best available designs, however, the probability of being able to detect policy-relevant effects is limited by how many schools the evaluation resources can serve.

Precision levels obtainable in the experimental-based evaluation strategy using schools as the unit of assignment are limited for two reasons: (1) the majority of students do not change breakfast program participation status (the "dilution" effect), and (2) variation exists between schools.

1. The "Dilution" Effect

The majority of students at the participating schools are unlikely to start receiving school breakfasts under the USBP. Many of them--on average 25 or 30 percent--will already be participating in the regular SBP, and it is expected that they will continue to receive breakfast under the USBP. The introduction of free school breakfasts is not likely to substantially affect their outcomes. Among the other students, a significant number will choose not to start eating a school breakfast, even if it is now free, because their families will continue to provide breakfast at home or for other reasons. The precision calculations are based on an assumption that the increase in school breakfast participation will be approximately 25 percentage points (from 30 to 55 percent). In the experimental versus control comparisons, which are the essence of the full application of the experimental design, the effects of school breakfast on those 25 percent that are new participants are diluted by the 75 percent of students whose participation is not changed.[1]

The design report outlines two possible approaches to dealing with this problem. The first is to conduct a preimplementation survey at the schools to identify the children most likely to

---

[1]Based on the information available, the 25 percent increase in participation represents a reasonable assumption. However, if FNS or others can supply information supporting a higher assumption, this would lower the estimates of minimum detectable effects. The effects are proportional--with a 50 percentage point rather than a 25 percentage point increase in participation, minimum effect sizes would be divided by 2. However, given that the program is likely to attract additional participants who previously did not choose to eat a school breakfast, many of whom may already be eating a nutritious breakfast at home, it is possible that program impacts may be smaller with large increases in participation. If so, even though the statistical precision of the evaluation design increases with larger increases in participation, the policy-relevant program impacts may be smaller than those reported currently in the literature.

participate and focus the analysis on those students. The second is to use analysis approaches that are not based directly on the experimental design. In particular, such nonexperimental methods would take advantage in the analysis of the fact that one can observe *which students actually participate in the USBP*, thus making it possible to sharpen the comparisons in the analysis (though at the cost of losing some advantages of the pure experimental design).[2]

2. Variation Between Schools

The other main problem that limits the precision in the currently planned demonstration is that, even with pairing of schools prior to randomization, there are likely to be significant differences between the schools. This variation tends to confound the analysis, because it is correlated with treatment status. For instance, if the treatment school in one of the school pairs happens to have children with relatively high nutrient intakes, the treatment group as a whole will reflect this, making it difficult to disentangle school differences from the effects of the USBP.

For a given number of schools, adding more children per school to the design has only limited effects in solving this problem. Even if data were available on *all* the children in a school and, hence, the evaluation could measure the school means perfectly, there would still only be a limited number of school observations with which to estimate treatment effects.

---

[2]Another possible approach to dealing with the dilution problem would be to require all SFAs participating in the demonstration to essentially enforce student participation by offering an intensive treatment in which all students would receive a breakfast in their classrooms, with no options for not participating. After careful consideration of this approach, this option is not recommended because of two significant research risks associated with it. First, it is unlikely that most of the SFAs applying for the program would agree to operate this variant of the USBP. Second, even if they *did* agree, many of the students forced to participate might throw their breakfasts away, thus, the intended effects of the intensive treatment and possibly leading to incorrect analysis inferences.

The evaluation design presented in this report addresses this main challenge to the evaluation, as well as the many other measurement and data collection challenges inherent to an evaluation of this magnitude.

C. OVERVIEW OF THE REPORT

The rest of the report is organized in four parts and provides a detailed strategy for evaluating the USBP pilot projects. The rest of Part 1 provides background information, describes the SBP, summarizes the literature on breakfast and learning, and discusses the rationale for the USBP. It also identifies the requirements of the demonstration and evaluation, provides a conceptual framework of how participation in a USBP may affect student outcomes, and provides an overview of the full set of evaluation design components.

Part 2 describes the design for the impact study. First, it identifies research questions to be addressed and the hypotheses to be tested; then, it discusses outcomes and the measures that will be used to assess program effectiveness. It describes the approach for evaluating the USBP demonstration programs, discusses evaluation and sampling design issues and strategies, presents an analysis plan for assessing program impacts, and describes data collection plans.

Part 3 presents the approach for conducting the implementation study of the USBP pilot projects. It describes study topics and research questions and outlines plans for process data collection. Also described are analysis strategies for assessing program implementation and whether the goals of administrative simplification are being met.

Part 4 provides an assessment of the feasibility of implementing the preferred evaluation design. It reviews the statistical precision achieved, provides an estimate of the demonstration costs, and identifies strengths and weaknesses, as well as risks.

II. BACKGROUND

In order to develop a design that results in an evaluation fully responsive to the needs of the Secretary of Agriculture and Congress, it is important to be clear about the demonstration and evaluation requirements specified in the legislation authorizing the Universal-Free School Breakfast Program (USBP) pilot projects. This chapter provides background information for the USBP demonstration and evaluation. Section A describes the School Breakfast Program (SBP), while Section B summarizes the literature on breakfast consumption and student learning. Section C provides the rationale for the USBP demonstration. Section D identifies and discusses demonstration and evaluation requirements. Section E describes a conceptual framework of how a universal-free school breakfast program would affect students and schools. Finally, Section F provides an overview of the evaluation components.

A. OVERVIEW OF THE SCHOOL BREAKFAST PROGRAM

The SBP is a federal program that provides states with cash assistance (and commodities) for breakfast programs in schools. It is administered at the federal level by the Food and Nutrition Service (FNS). State education agencies and local school food authorities (SFAs) administer the program at the local level. Originally, it was a pilot program that targeted children from low-income school districts. With the 1975 amendments to the Child Nutrition Act of 1966, the SBP became permanent. The objective of Congress was to make the program available in all schools where it was needed, to provide adequate nutrition for children in attendance. Congress passed the 1989 Child Nutrition Act to expand the availability of the SBP, requiring the Secretary of Agriculture to award start-up funding (ranging from $3 to $5 million per school district) to states wishing to implement

the SBP in schools with a large proportion of children from low-income households (Kennedy and Davis 1998).

All public and (nonprofit) private elementary and secondary schools in the United States are eligible to participate in the SBP. To participate, schools must make breakfasts available to all students, and the breakfasts must meet federal nutrition standards. SBP breakfasts are required to provide approximately one-fourth of the Recommended Dietary Allowance (RDA) for food energy and other important nutrients over a period of time (protein, calcium, iron, vitamin A, and vitamin C). In addition, regulations now require that all school meals meet the recommendations of the *1995 Dietary Guidelines for Americans* (U.S. Department of Agriculture and U.S. Department of Health and Human Services 1995).[1]

To achieve both the RDA and the *Dietary Guidelines* standards, schools may use several methods for planning menus. One way is to prepare meals using food-based menu planning.[2] A school breakfast using the food-based menu planning approach must contain, at a minimum, the following food components:

- A serving of fluid (whole or low-fat) milk served as a beverage or on cereal or used in part for each purpose
- A serving of fruit or vegetable or both, or undiluted fruit or vegetable juice
- Two servings from one of the following components--bread/bread alternate or meat/meat alternate. Alternatively, there can be one serving from each component.

---

[1]The applicable recommendations of the *Dietary Guidelines* are to (1) eat a variety of foods; (2) limit total fat to 30 percent of calories; (3) limit saturated fat to less than 10 percent of calories; (4) choose a diet low in cholesterol; (5) choose a diet with plenty of vegetables, fruits, and grain products; (6) use salt and sodium in moderation; and (7) eat a diet rich in dietary fiber.

[2]The other two methods are nutrient standard menu planning and assisted nutrient standard menu planning.

The USDA subsidizes schools for each eligible breakfast served. Schools submit a claim to their SFA, which submits a claim to the state agency; the state agency then submits claims to the USDA. The USDA reimburses the state, which in turn reimburses the local SFA, and then schools. The cash reimbursements vary according to whether students qualify for free, reduced-price, or full-price meals. To be eligible for free meals, students must have a family income less than or equal to 130 percent of the poverty level. To be eligible for reduced-price meals, students must have a family income between 130 and 185 percent of the poverty level. Those with family income levels greater than 185 percent of poverty pay full price (which is actually less than market price). For the 1999-2000 school year, the reimbursement in "non-severe-needs" schools (schools that do not have a large proportion of needy individuals) in the contiguous United States is $1.09 for free breakfasts, $0.79 for reduced-price breakfasts, and $0.21 for full-price breakfasts. For schools with a large proportion of needy individuals ("severe-needs" schools), reimbursements are $0.20 higher for free and reduced-price breakfasts.[3]

SBP participation grew rapidly from 1970 to 1980, but more modestly through the 1980s (Kennedy and Davis 1998). Participation has grown dramatically over the past decade. The number of schools offering the SBP increased from 46,100 in fiscal year 1991 to 68,718 in fiscal year 1997, an increase of almost 50 percent. It is estimated that nearly three-quarters of schools that offer a school lunch also offer breakfast (Marcotte 1999).

B. EFFECTS OF BREAKFAST CONSUMPTION ON CHILDREN'S LEARNING: A SUMMARY OF THE LITERATURE

Existing literature suggests a relationship between eating breakfast on the one hand and improved dietary status and enhanced cognitive performance, on the other. Although the studies

---

[3]Reimbursement rates for all meals are higher in Alaska and Hawaii.

suggest positive educational benefits, none of the studies has been able to conclude definitively that *participation* in the SBP (or USBP) causes improvements in either long- or short-term cognition and school performance. The inconclusive findings reflect limitations in the studies undertaken, such as using unreliable measures of breakfast participation and school performance or the use of nonexperimental pre-post or comparison group designs that do not adequately control for selection effects.

The rest of this section summarizes the literature. It is organized around three issues: (1) the link between nutrition and cognitive development of children; (2) the contribution of breakfast to children's dietary intake and behavioral and cognitive development; and (3) the relationship between school breakfasts, dietary status, cognition, and achievement.

1. Nutrition and Cognitive Development

There is an extensive literature documenting the long-term effects of inadequate nutrient intakes in developing countries and a more limited literature on developed countries such as the United States. Early research focused on severe protein-energy malnutrition. Permanent structural damage to the brain has been attributed to the extremely detrimental effects of severe undernutrition (Brown and Pollitt 1996)--a level of malnutrition that is almost never seen in the United States.

The literature on developed countries has focused primarily on the effects of mild to moderate undernutrition. In the United States, mild to moderate undernutrition is most often seen among infants and young children, in the form of low birth weight, intrauterine growth retardation (small size for gestational age), or "failure to thrive" in the period after birth. Although these problems are not confined to poor families, they are strongly associated with poverty. In addition, low height for age (stunting) and low weight for height (wasting) are more prevalent in the United States among

poor children than nonpoor children, and they are particularly prevalent among those who are persistently poor (Miller and Korenman 1994).

Much of the early research attributed the detrimental effects of undernutrition to permanent structural damage to the brain during the critical growth periods of pregnancy and the first year of life, when brain growth is most rapid (Brown and Pollitt 1996). There is considerable evidence of long-term effects of low birth weight on cognitive development; lower birth weights are associated with more severe cognitive impairments (Hack et al. 1995). Recent studies, however, suggest that important aspects of cognitive development occur both before and after periods of rapid brain growth, suggesting that neurological damage from undernutrition may occur at other times as well.

Other research explores the hypothesis of "functional isolation" (Levitsky 1979), which posits that undernourishment leads to poor motor development and lower activity levels and responsiveness. Lower activity levels, in turn, reduce young children's exploration of their environment and the stimulation they receive from caretakers; as a result, they experience developmental delays. The evidence is mixed for the hypothesis that reduced activity levels are a mediating factor in cognitive development (Meeks et al. 1995). Regardless of the cause, there is evidence that stunting has effects on motor and cognitive development (Wachs 1995; Simeon and Grantham-McGregor 1990; and Pollitt et al. 1996).

Recent research also indicates that undernutrition during any period of childhood, even for relatively short episodes, can have negative effects on cognitive development (Center on Hunger, Poverty, and Nutrition Policy 1994). Studies of the effects of nutritional supplementation programs, however, suggest that these programs may ameliorate the effects of nutritional deficits, even if the interventions occur after the early periods of rapid brain growth (Pollitt et al. 1996).

Recent studies highlight the relationship between hunger and psychosocial functioning of children. Researchers from the Community Childhood Hunger Identification Project (CCHIP) found that children who were either hungry or at risk of being hungry were twice as likely as nonhungry children to be classified as having impaired psychosocial functioning (Murphy et al. 1998(a)). At school, hungry children are more likely than nonhungry children to be irritable, anxious, and aggressive; they also are more likely to be absent from or late to school.

2. Eating Breakfast, Daily Nutrient Intake, and Behavioral and Cognitive Development

Studies that focus on the role of breakfast typically examine its contribution to daily nutrient intakes or its relationship to the likelihood of satisfying dietary standards. Although these studies vary greatly in the study populations and data sets used, a consistent finding of the studies is that breakfast makes a significant contribution to nutrient intake over 24 hours. For children, analysis of data from the first National Evaluation of the School Nutrition Program showed that eating breakfast was significantly and positively related to the daily intake of all nutrients examined (Devaney and Fraker 1989). Later studies of breakfast consumption patterns of children also found that total daily intakes of food energy and other nutrients were significantly lower for children who did not consume breakfast, compared with children who did consume breakfast, and that children who skipped breakfast did not make up the difference in nutrient intakes at other meals (Nicklas et al. 1993; and Sampson et al. 1995). Similar findings are reported by other studies of breakfast consumption, most notably those based on data from the Bogalusa Heart Study. These studies examined trends in eating breakfast and the nutrient contribution of breakfast, and concluded that breakfasts, especially those containing ready-to-eat cereals, have a significant impact on the nutritional quality of diets of children and young adults (Nicklas et al. 1998).

Evidence on the effects of eating breakfast generally indicates improved short-term cognitive performance. A thoughtful review of studies of breakfast consumption concluded that eating breakfast appears to improve performance on specific cognitive tasks involving memory, although significant shortcomings of many studies are noted (Pollitt and Mathews 1998). The effects of eating breakfast on performance on cognitive tests appear even more pronounced after a period of fasting and for more vulnerable subgroups of children, such as those that are nutritionally at risk (Simeon and Grantham-McGregor 1989; and Pollitt et al. 1998). Long-term assessments of breakfast omission and cognitive function have not been conducted (Pollitt and Mathews 1998).

3. School Breakfasts, Dietary Status, and School Performance and Achievement

Several studies have been done on the effects of the SBP on students' dietary intake; these studies have found that the SBP has been successful in improving the nutrient intake of program participants, particularly of low-income children. Some studies have examined the effects of breakfast program participation on school-related outcomes such as student achievement, cognition, attendance, and psychosocial measures. These studies have generally found that participation is linked with higher attendance; less is known about the effects of the breakfast program on students' cognition and academic achievement. Moreover, the studies have suffered from various methodological limitations, so the question of how a USBP would influence children's school-related outcomes remains open.

a. Effects of School Breakfasts on Students' Dietary Status

School breakfasts must be substantial to supply significant amounts of nutrients to children. Regulation requires that school breakfasts provide approximately one-fourth of the RDA for important nutrients. A study of meals offered by the SBP during spring 1992 found that most school

breakfasts were relatively simple, typically made up of a few choices from the bread or bread alternate food group and a choice from the fruit, vegetable, or juice food groups (Burghardt et al. 1995). School breakfasts rely heavily on breads and ready-to-eat cereals.

Studies of dietary intake of students indicate that the SBP is generally successful at meeting the program goal of providing a nutritious breakfast to children who might not otherwise receive one. A recent study of breakfast consumption found that the percentage of students eating a "robust" breakfast (intake of food energy greater than 10 percent of the RDA) is higher for low-income students attending schools with the SBP than for similar students attending schools without the SBP (Devaney and Stuart 1998). In addition, findings from the first School Nutrition Dietary Assessment Study (SNDA-1) showed that SBP participants had significantly higher intakes of food energy, protein, thiamin, riboflavin, phosphorus, magnesium, and calcium than nonparticipants (Devaney et al. 1995; and Ponza et al. 1997). Other studies also document increased intakes of key nutrients for students eating school breakfasts (Devaney and Fraker 1989).

b. Effects of School Breakfast on Attendance and Tardiness

USBP participation is associated with higher rates of attendance (Abell Foundation 1998; Cook et al. 1996; Jacoby et al. 1996; Meyers et al. 1989; Murphy et al. 1998a, 1999; and Powell et al. 1998). Several studies also found that participation is associated with declines in tardiness (Abell Foundation 1998; Cook et al. 1996; Meyers et al. 1989; Murphy et al. 1998b, and Murphy et al. 1999). The size of these effects is moderate, and some of the studies either did not conduct significance tests or had relatively small samples and did not find the effects to be statistically significant. However, since the findings are shared by many studies, based on different samples and using different methodologies, the overall finding that breakfast program participation leads to

higher attendance and less tardiness is credible. This conclusion is shared by the most recent major review of this literature prior to the current project (Pollitt and Mathews 1998).

c. Effects of School Breakfast on Cognition and Academic Achievement

The estimated effects of participation in school breakfast programs on academic achievement have been mixed in previous studies. Meyers et al. (1989) found the largest effects, with participation in the regular SBP estimated to lead to a significant increase of 10 percent of a standard deviation in a child's battery score on the Comprehensive Test of Basic Skills (CTBS). Even this study, however, failed to find statistically significant effects of participation on the subtests that make up the CTBS: participation was estimated to have positive, but not significant, effects on language and math subtest scores, and essentially no effect on the reading subtest score.

Two other studies of USBP programs in the United States are relevant. Murphy et al. (1998a) did not examine test scores but found that USBP participation in two large eastern cities was positively and significantly related to students' math grades. In addition, Wahlstrom et al. (1997) presented data on mean test scores before and after USBP implementation in several USBP and comparison schools in Minnesota. However, this study conducted no significance tests and made no claims about the implications of these data with respect to the effects of the USBP on academic achievement.[4]

Finally, a few studies have also examined the effects of breakfast program participation on other, related outcomes. Murphy et al. (1999) and Wahlstrom et al. (1997) found that being in a USBP school was associated with decreases in nurse visits and improvements in teacher and parent

---

[4]Other studies (Powell et al. 1998; Chandler et al. 1995, and Jacoby et al. 1996) also have examined the effects of breakfast program participation on student academic achievement. However, because these studies are of students in developing countries, they are of limited relevance to U.S. students.

perceptions of the learning environment in school (although these relationships were not statistically significant). Murphy et al. (1998b) and Murphy et al. (1999) found that USBP participation was significantly associated with children's psychosocial outcomes, arguing that the program led to lower levels of such symptoms as anxiety, hyperactivity, childhood depression, and psychosocial dysfunction.

d. Methodological Issues

Several design and methodological issues are important for understanding the existing evidence on the effects of school breakfasts on school outcomes and planning for a rigorous, definitive evaluation of the USBP. These issues are:

- ***Limited Attention Devoted to Any One Outcome.*** Only a relatively small number of studies have examined the effects of school breakfasts on any given outcome (except for attendance/tardiness).
- ***Differences in the Breakfast Program Interventions.*** Previous studies have examined different types of breakfast programs serving specific populations. Their applicability to implementing a USBP is uncertain.
- ***Nonexperimental Designs.*** Most of the studies used nonexperimental designs that were potentially subject to selection bias.
- ***Small Sample Sizes and Inappropriate Statistical Tests of Significance.*** Existing studies often analyzed small samples of both students and schools, and most did not adequately account for their clustered samples in their significance testing.

Each of these issues is discussed below.

Relatively Small Number of Studies Have Been Conducted on School Outcomes. Compared with the number of studies examining the effects of eating breakfast on behavioral and cognitive development or the general link between nutrition and cognitive development, a relatively small number of studies examined the effects of *school breakfasts* on these outcomes. For example, only two studies of the SBP examined how participation influenced achievement test scores, and no

studies of the SBP focused on short-term cognitive outcomes. Similarly, two studies examined students' psychosocial outcomes, and another two focused on students' visits to the school nurse. Given some of the methodological limitations of these studies, the fact that they are also few in number makes drawing definitive conclusions from them difficult.

Differences in the Program Interventions. An important consideration is that breakfast feeding programs (and their evaluations) vary considerably in the populations served and the intervention provided. Many of the existing studies focus on breakfasts provided to children in developing countries (Chandler et al. 1995; Jacoby et al. 1996; and Powell et al. 1998). Although these studies provide useful information, one cannot necessarily infer the effects of participation in the SBP on the basis of the estimated effects of the Jamaican or Peruvian school breakfast programs.

Even those studies of the SBP vary from what might be expected from a USBP. One of the strongest studies--the study of the impacts of introducing a SBP in Lawrence, Massachusetts (Meyers et al. 1989)--compared outcomes among low-income students after the SBP program was implemented with outcomes before the SBP. It is not clear whether the effects of participating in a USBP, which is most likely to be implemented in a school already operating the SBP, would be the same as the effects of participating in the regular SBP.

Nonexperimental Design. Existing studies of the SBP have used nonexperimental designs: individual students (or schools) chose on their own whether or not to participate in the program. Thus, there was no guarantee that program participants were similar to nonparticipants, and this design necessitated controlling for relevant preexisting differences between the two groups when measuring differences between the groups in the outcome measures. The studies controlled for the preexisting differences in a variety of ways, although most used some sort of pre-post design among both treatment group and control group members. However, each of these studies is subject to the

potential criticism that its findings are driven more by the preexisting differences between participants and nonparticipants than by the effects of the breakfast program. The internal validity of these studies is not as great as the internal validity of the experimental studies.

Small Sample Sizes. Most of the studies used relatively small samples. For example, Murphy et al. (1998b) analyzed a sample of 133 students in three schools, and the Abell Foundation (1998) analyzed schoolwide data from just three USBP and three non-USBP schools. Even in studies that used large samples of students, these students came from relatively few schools. For example, Meyers et al. (1989) analyzed a sample of more than 1,000 students, but these students came from just six schools within a single school district.

In principle, tests of statistical significance allow researchers to determine the extent to which they can be confident that their estimates reflect true effects of participation rather than random chance, and these significance tests take into account the sample sizes of students. However, when samples of students are clustered in a small number of schools, as is true for all of these studies, the statistical tests of significance must take the cluster effects into account. Although none of these studies states how exactly its statistical tests of significance were conducted, it does not appear that they accounted for school-level variance in their analyses. Given the relatively small samples of schools in these studies, properly taking into account this correlation across different sample members within the same school would likely have led to dramatically lower significance levels.

C. RATIONALE FOR THE USBP PILOT PROJECTS

Despite the increase in the number of schools offering the SBP, the percentage of students eating school breakfasts is considerably lower than the comparable percentage eating school lunches. Moreover, compared with students eating school lunches, those eating school breakfasts are more likely to be poor and to qualify for free or reduced-price meals. One reason why participation in the

SBP is lower than in the School Lunch Program may be the timing of the meal: breakfast is typically served prior to the start of school, whereas lunch is provided during school hours. Because students who participate in the SBP are more likely to be low-income, another, possibly more important reason that there is reduced participation in the SBP, may be students' perceived stigma associated with the use of free and reduced-price school meals.

One approach to increasing participation in the SBP is to offer breakfast free to all students, regardless of their ability to pay for meals. It is thought that a "universal-free" school breakfast program removes the stigma often associated with school breakfast, resulting in more children (both poor and nonpoor) participating. It is believed that a universal-free program will result in more children consuming nutritious breakfasts, thus allowing students to begin school with the proper nourishment needed to learn. Better nourishment in the morning will translate into better school performance throughout the rest of the day, but especially during mid- to late morning.

Expanding the SBP so that breakfasts are free to all students could, however, substantially increase the cost to the federal government of subsidizing school breakfasts, should the USBP encourage participation to the extent that its proponents believe it would. In a climate where public resources are constrained, it is critical to know whether these expenditures are worthwhile. Key issues are whether (1) increased participation in the SBP results in improved dietary intake, academic performance, and related behaviors; and (2) whether the USBP meals are simply substituting for meals that students, particularly students from nonpoor households, would otherwise eat in the absence of the USBP.

Because of the absence of hard evidence on the effects of the USBP on student outcomes and costs, the 1998 Child Nutrition Reauthorization Act called for a demonstration to evaluate the effects of providing free breakfasts to elementary school children. The Act states that the Secretary shall

make grants to state agencies to conduct pilot projects in elementary schools under the jurisdiction of not more than six school food authorities approved by the Secretary, to:

- Reduce paperwork, simplify meal counting requirements, and make changes that will increase participation in the school breakfast program
- Evaluate the effect of providing free breakfasts to elementary school children, without regard to family income, on participation, academic achievement, attendance and tardiness, and dietary intake over the course of a day

The extent to which a universal-free breakfast both increases participation and breakfast consumption and improves children's dietary and cognitive outcomes, and the implications for program costs, need to be documented. The objective of the demonstration and its evaluation is to assess the impacts of the USBP. The purpose of this report is to develop an evaluation design that will yield valid results.

D. DEMONSTRATION AND EVALUATION REQUIREMENTS

The 1998 Child Nutrition Reauthorization Act sets a number of guidelines for the demonstration and evaluation. Key features of the demonstration that the design must take into account, as specified in the legislation and interpreted by FNS, are as follows:

- ***Eligible Schools.*** The demonstration is restricted to elementary schools--kindergarten through grade 6.
- ***Number of School Food Authorities.*** The demonstration is to be conducted in elementary schools in not more than six School Food Authorities (SFAs).
- ***Selection of SFAs.*** The demonstration is voluntary. The USDA will invite school districts to compete for the pilot projects by submitting applications. The Federal Register Notice announcing the demonstration was released in early December, 1999.[5] SFAs wishing to participate will request an application and submit it to their state

[5]See *Federal Register*, vol. 64, no. 233, Monday, December 6, pp. 68077-78.

administrative agency by January 31, 2000. The application will request information on the characteristics of schools in the SFAs. The state agency will submit nominations to FNS of SFAs wishing to participate. These nominations must be sent to FNS by February 15, 2000. FNS will form a selection committee. FNS is planning on selecting the sites by April 1, 2000. SFAs will be selected so as to obtain representation of pilot projects among urban and rural elementary schools, and among elementary schools of varying family income levels. To permit a valid evaluation of program effects, SFAs will need to be sufficiently large to yield an adequate number of elementary schools that are participating and not participating in the USBP. Given the small number of SFAs to be included (six SFAs) and the size requirements, the SFAs will need to be purposively sampled.

- ***Length of the Demonstration.*** The demonstration is to last three (successive) school years.

- ***Demonstration Startup.*** While not specified in the legislation, in order to meet reporting requirements for the evaluation's interim report (see below), it is desirable that the demonstration begin in school year 2000-2001 (either fall 2000 or spring 2001). It would run through school year 2002-2003.

- ***Emphasis on Internal Validity.*** The objectives of the evaluation are to assess the impact of the USBP on student participation in the breakfast program and student outcomes, and on the costs to the federal government in terms of increases in funding for reimbursing program meals, as well as how it simplifies paperwork required to be completed by the schools and other administrative procedural participation requirements. Reflecting the limited number and purposive selection of SFAs, FNS has indicated that priority should be placed on the evaluation objective of obtaining reliable and defensible findings as to whether the USBP increases participation and improves student outcomes in pilot project schools (internal validity), over the ability to generalize results to other settings (external validity).

- ***Timing of Evaluation Reports.*** At least two reports summarizing evaluation findings are to be submitted. The final version of the evaluation's interim evaluation report will need to be submitted in spring 2002, to inform the Act's reauthorization hearings. No date was given for submitting the final report; however, it is likely that the final version of the evaluation's final evaluation report will need to be submitted when the present Act expires, at the end of September 2003.

- ***Funding for the Demonstration and Evaluation.*** Congress appropriated $7 million for the evaluation and demonstration in USDA's fiscal year 2000 budget. FNS expects that, ultimately, approximately $13 million will be available to conduct the USBP demonstration and evaluation.

1. Discussion

Some possible limitations associated with the current plans for the demonstration should be noted. They are discussed below.

a. Limitation to Six SFAs

The limitation of the demonstration to no more than six SFAs unnecessarily limits the evaluation. Relaxing the six-SFA constraint to permit the evaluation to include more districts could make it easier to achieve representation of rural districts and schools without sacrificing statistical precision. Significantly, this could be accomplished without exceeding the $13 million funding constraint.[6] However, it is unlikely that the 6-SFA constraint will be changed.

b. Start Date

Starting the demonstration in fall 2000 may be difficult to achieve. FNS will need to release an RFP and award an evaluation contract, secure OMB clearance, and recruit school districts for the demonstration, and districts will need to secure approval from Boards of Education--all within the next nine months. Moreover, to provide an opportunity to conduct baseline data collection from

---

[6]The six-SFA constraint was included primarily as a way to limit demonstration costs. Demonstration costs, however, are driven largely by the number of schools, not SFAs. To be sure, holding the number of schools in the demonstration constant, adding SFAs would increase the costs of the evaluation, because there are fixed costs to the evaluation contractor of coordinating with the SFA/school district to conduct random assignment and retrieve school records data, as well as the costs of conducting the SFA director interviews. However, these costs are substantially lower than demonstration (meal subsidy) costs for participating USBP demonstration schools. Ultimately, the demonstration costs depend on the size of the increase in participation for the school breakfast program resulting from the introduction of the USBP. These costs are estimated to be between $21,000 and $32,000 per USBP school per year, assuming 490 students per school, an initial SBP participation rate of 0.30, and 180 meal serving days. The low estimate is based on participation increasing 15 percentage points (from .30 to .45) and the higher estimate based on participation increasing 25 percentage points (from .30 to .55). Thus, demonstration meal costs for each USBP school are expected to be between about $60,000 and $100,000 over the three-year demonstration.

students in fall 2000 (in order to conduct a preimplementation student survey), it will be desirable for program implementation to begin somewhat later, say, in January 2001.

2. Selecting SFAs

When selecting school districts and schools to be included in the demonstration, one would ideally want to stratify school districts by characteristics such as region, size of district, urbanicity, and schools containing students with varying family income levels, then randomly select SFAs from each stratum. This approach, however, is not possible, given that, at most, six SFAs are to be included in the demonstration. Instead, districts will need to be purposively selected.

Further, the calculations in Chapter V suggest that, to reliably estimate student outcomes, the evaluation would need to include at least 120 schools. This implies that the evaluation will need to include the largest districts. The evaluation would need to purposively select one, or perhaps two, smaller "rural" districts to ensure representation of rural schools. All else the same, this approach would reduce precision somewhat. The evaluation could compensate for this by including a couple of larger districts, or it may be the case that the demonstration will be able to select a large countywide school district that offers SBP to a number of rural schools within the county, if such a county is nominated.[7]

There are other factors, in addition to those already mentioned, that FNS may want to consider when selecting SFAs to participate in the demonstration. For example, FNS may want to consider evaluation design criteria such as those that could make data collection easier and cheaper:

---

[7]"Pooling" districts--grouping smaller rural districts into larger, single districts to achieve rural representation and at the same time reach the target number of schools--was considered. However, the legislation is fairly specific that the demonstration be conducted in no more than six SFAs, so that is not a viable option.

- ***Regular Student Testing.*** Districts that have annual testing in every grade would be ideal for learning about the effects of USBP on student outcomes. It would also help if such tests were based on, or equated to, nationally administered and nationally normalized tests like the California Achievement Test (CAT), the Iowa Test of Basic Skills (ITBS), the Stanford Achievement Test, or the Comprehensive Test of Basic Skills (CTBS). If the option of annual testing in every grade is not available, the next best one would be testing in adjacent grades, such as grades 1 and 2 or 4 and 5. The adjacent-grade tests would not have to be in the same subject, although similar subjects are preferable.

- ***Satisfactory Management Information System.*** Districts that can easily and cleanly retrieve data on students at the individual level, as well as data on specific schools, would contribute more to the evaluation than districts with less effective information systems, and would reduce the burden on schools, teachers, and parents to provide important data elements. Factors to take into consideration may include the comprehensiveness of the student-level records (detailed and accurate information on students and their families), consistency of reporting formats across schools and over time, and ability to mask individuals' identity to protect confidentiality.

- ***Having Many Elementary Schools in the District with Common Grade Configurations.*** Districts with few schools or with schools that have different grade configurations (K-2, K-4, 3-6) could drive up the costs of the USBP evaluation, because of added complexity. A much simpler study would include only districts with elementary schools housing all primary grades in the same school--grades K-5, K-6, and so on. Because the demonstration is limited to six districts, it would increase precision to have as many schools per district as possible, without overly skewing the sample of districts.

- ***High Desire to Be Actively Supportive to the Evaluation.*** It will be important for the selected SFAs and schools to support the evaluation team in providing high-quality data.

3. Required SFA Application Data

In order to address the above issues, it is recommended that the application collect information on:

- How many elementary schools are in the district?
- What are the grade configurations of the elementary schools in the SFAs?
- How many elementary schools are in rural, suburban, and urban locations?
- What is the total student enrollment in the district's elementary schools?

- How are breakfasts currently served under the regular SBP? How does the district intend to operate the universal-free breakfast program?

- Does the school district administer standardized achievement tests to elementary school students? Which tests does it administer?

- What districtwide test(s), if any, will definitely be in use for grades 1 through 6 in each of the following school years: 1999-2000; 2000-2001; 2001-2002; 2002-2003?

- Does the district use adaptive testing? (Adaptive tests customize the difficulty of test items to each student's estimated test performance level.)

- For each possible data element (for example, attendance, absences due to illness, tardiness, disciplinary incidents, school breakfast/lunch eligibility, special education status, Limited English Proficient status, race or ethnicity, family income, household composition, other), would the data be available for a demonstration? If so, does the district maintain individual student records and/or data aggregated to the school level? Finally, are historical data available?

- Is public school assignment determined by residence only? If no, how is school assignment determined (for example, open enrollment areas, magnet schools, charter schools, other school choice)?

- Approximately what percentage of students in a typical elementary school in the district leave their school between the beginning of one school year and the beginning of the next school year (student mobility)?

- Would the school district be willing to work with the evaluation contractor to conduct a preimplementation survey of students' parents (e.g., release the names, addresses, and telephone numbers of parents of children in the schools)?

E. CONCEPTUAL FRAMEWORK

The demonstration will offer free breakfasts to all schoolchildren in the USBP schools, under the assumption that eating school breakfasts will have short- and long-term beneficial impacts. The underlying rationale is as follows:

- Some children either do not eat breakfast or do not eat a nutritious breakfast.

- Providing free school breakfasts will increase the number of school breakfasts consumed and thereby increase the number of students who eat a nutritious breakfast.

- Eating a nutritious school breakfast leads to enhanced readiness to learn, improved cognitive and behavioral outcomes, better dietary status, and, ultimately, higher academic achievement and school performance.

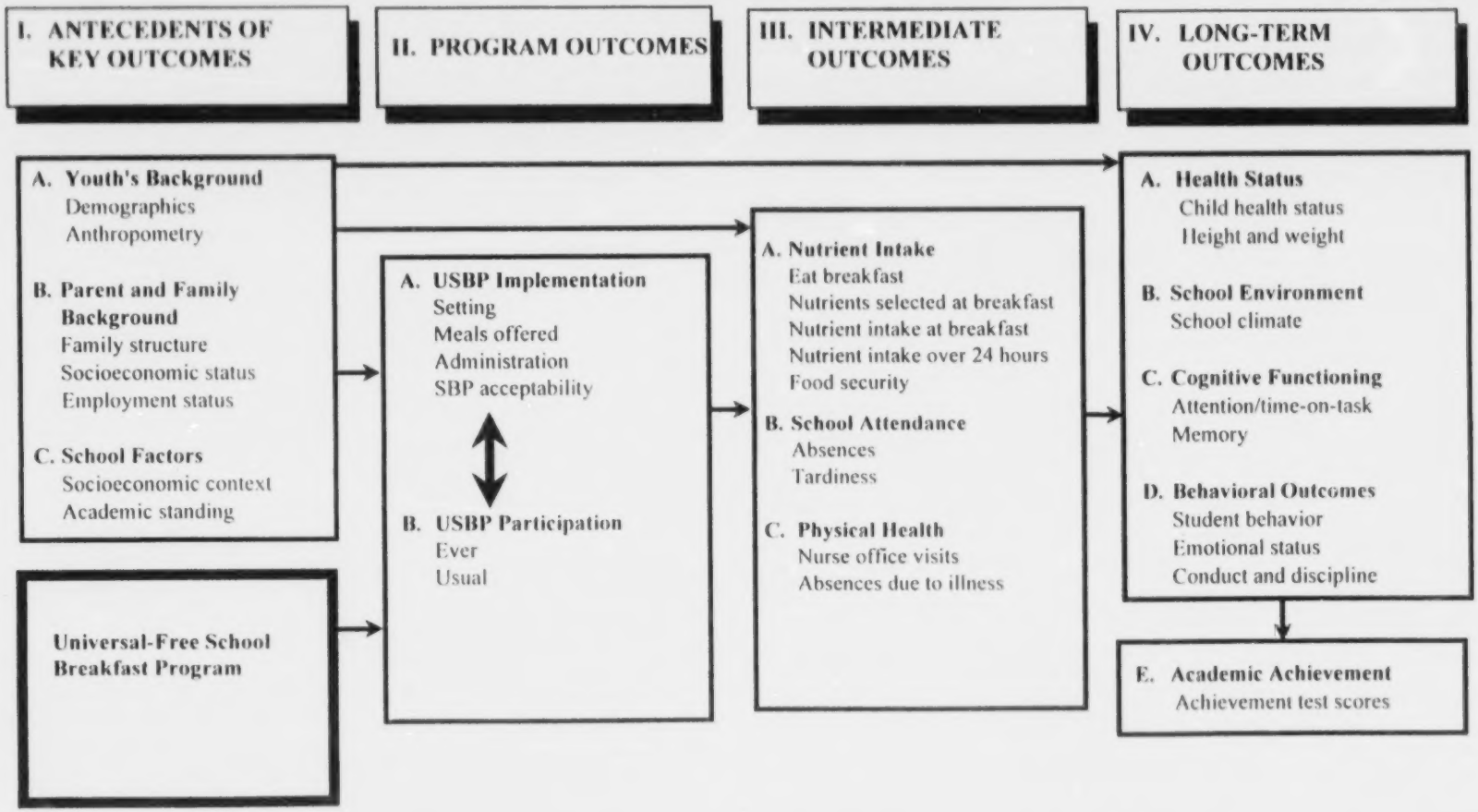
Other factors influence these outcomes, but the USBP could be a major external factor affecting whether schoolchildren eat breakfast, what kind of breakfast they eat, and how they perform in school.

Figure II.1 suggests a conceptual framework of the possible effects of the USBP. This framework highlights the important link between the background of children--including family and school influences; program outcomes resulting from the introduction of the USBP; and intermediate outcomes relating to nutrient intake, school attendance, and health--and resulting changes in cognitive and behavioral outcomes and academic achievement. The framework offers a starting point for understanding how and why the USBP affects program participation and outcomes, while also helping frame the later discussion of evaluation design alternatives and key measurement and analysis issues.

The antecedents of key outcomes (Column I) are primarily the background characteristics of the students and their families. These antecedents include such individual characteristics as ability, health, age, gender, race and ethnicity; and anthropometric measures such as height and weight. They also include social and demographic characteristics of the family (such as socioeconomic status), as well as some aspects of the school context and environment. Any of the background factors may have an important, direct effect on the long-term outcomes, or they may operate indirectly by influencing either program participation or intermediate outcomes, as shown in Columns II and III. For example, independent of any programs available, these background factors may have strong influences on the usual nutrient intake, which in turn affects key long-term cognitive, behavioral, and academic achievement outcomes.

FIGURE II.1

CONCEPTUAL FRAMEWORK FOR ANALYZING THE EFFECTS OF A UNIVERSAL-FREE SCHOOL BREAKFAST PROGRAM



For the USBP demonstration and evaluation, the focus is on determining how the introduction of a universal-free breakfast program alters, first, whether students eat nutritious breakfasts and, second, the effects of these breakfasts. The program may affect participation in two ways. First, because breakfasts are now provided free, children who previously ate breakfast at home or who do not usually eat breakfast at all may eat breakfast at school. Second, various changes in program operations may influence whether students participate. For example, the USBP may reduce any stigma associated with eating school breakfasts. Or schools may choose to incorporate the meal program into the student's classroom (such as by serving the breakfast during class time rather than before school), as opposed to serving it in the cafeteria, thus making it more accessible to students. Also, if participation increases sufficiently, the scale of program operations may change, resulting in changes or improvements to the meals offered that could increase the perceived quality of school breakfasts and lead to increased participation.

Participation in the USBP may influence longer-term cognitive, behavioral, health, and school environmental outcomes--and, ultimately, through intermediate outcomes (Column III), school achievement (Column IV). For example, if USBP students regularly eat nutritious breakfasts at school, and these breakfasts are more nutritious than what students would get otherwise, they may have better academic outcomes resulting from long-term breakfast consumption patterns that enhance cognitive functioning and learning. Students who participate regularly in the USBP may be more likely to attend school, be tardy less often, and have fewer nurse visits due to inadequate breakfasts, thus enabling them to gain more exposure to school instructional time and spend more time on tasks. These changes in behavior may influence longer-term academic performance.

This conceptual framework suggests several issues and research questions for the evaluation design. First, does a universal-free breakfast increase student participation in the school breakfast

program, and are children more likely to consume breakfast? Second, does participation improve the nutrient intake of students in the morning and improve key cognitive, behavioral, and academic outcomes? Third, can the effects of the USBP on these outcomes be explained through its effects on program participation, other changes to the school breakfast program operations, and intermediate outcomes? For example, how do changes in participation affect nutrient intake at breakfast, and how does nutrient intake at breakfast contribute to cognitive functioning? Finally, do the effects of the USBP vary with any of the background characteristics of students and their families?

The next section explores the implications of this conceptual framework for developing evaluation design options for the USBP demonstration and evaluation.

F. OVERVIEW OF THE USBP EVALUATION

Building on the conceptual framework, an evaluation strategy for the USBP demonstration should be structured to provide evidence of the effects on a broad range of outcomes, including program participation, nutrient intake, cognitive and behavioral outcomes, student achievement, other school outcomes, health, and program operations and costs. To measure and examine these outcomes, the evaluation design approach presented in this report include both a comprehensive implementation analysis and an impact analysis. Together, these two broad evaluation components will provide:

- Descriptive information about USBP participants, SBP participants, and nonparticipants
- Analysis of the impacts of the demonstration on program participation and student outcomes
- Information on how the USBP was implemented, its cost, and any changes to program operations

The discussion below provides an overview, thus creating a context for what comes later. Detailed plans are discussed in later chapters.

1. Implementation Analysis

The implementation analysis will document how the USBP is implemented, how it differs from the regular SBP, and whether there are any changes to the scale or other aspects of program operations resulting from a universal-free school breakfast program. It will also examine whether there are changes in paperwork for SFAs and schools, and the effects on meal subsidy costs. The implementation analysis (Figure II.1) will focus on the program outcomes, especially those listed under USBP implementation. Topics to be covered will include the setting in which the USBP breakfasts are offered (that is, cafeteria, classroom, or other location), the types of food and nutrients in the meals offered under the USBP, plate waste (how much of the various nutrients in USBP breakfasts students waste), administrative changes, and acceptability of the USBP. Data for the implementation analysis will be obtained from (1) semistructured interviews with SFA staff, the school principal or school coordinator, and the cafeteria manager; (2) surveys of meals offered and menus; (3) interviews with students to compare school breakfast food items selected versus consumed; and (4) SFA and school records.

2. Impact Analysis

The design of the impact analysis for the USBP has several core features:

- The design includes a rigorous evaluation methodology, sampling plan, and analysis plan that are highly defensible, can withstand intense methodological scrutiny, and will provide the strongest possible evidence on the effects of the USBP. Reflecting the fact that some components of the design cannot be finalized until information is obtained on the SFAs applying and selected for the demonstration, the proposed impact study design includes options.

- It includes a broad range of outcomes to be examined. Although improved academic achievement and child health status may be the ultimate goals of a nutritious breakfast program, it also is important to look at short-term markers of success, such as changes in participation, nutrient intake, cognition, and other school-related outcomes.

- The data collection plan calls for longitudinal data collection in order to track outcomes over time, as well as to determine the impact of the USBP on academic achievement.

The general strategy for the impact analysis is to randomly assign schools to USBP or control status. Differences between student outcomes for demonstration versus control status will then be examined. The aim of the analysis is to estimate the effects of USBP availability on student outcomes. A major issue to consider is that the estimated effects will be based on samples drawn from all students attending the selected schools. Some of these students will choose to eat a school breakfast; others will not. Since the USBP is unlikely to affect outcomes among students who do not eat school breakfasts, the availability analysis should be interpreted only as providing information on the effects of the USBP on a group of students who are *offered* a free breakfast. It is possible that the USBP may have a strong effect on those who eat school breakfasts, but its overall effect will depend both on the extent to which students in the schools participate and the impact on those who participate.

If, as expected, the impacts on student outcomes that are generated by the introduction of a USBP operate solely through changes in outcomes *of new school breakfast participants* (that is, students who do not participate in the regular SBP program but choose to participate in a USBP), then impact estimates based on samples of the entire student population will be diluted, and therefore difficult to detect. For example, if the impact on an outcome measure is "X" for new participants, and the proportion of students who are new participants is .25 (as expected), the impact over the

whole student population will be .25X.[8] The challenge for the impact study design is to develop an approach that will enable the evaluation to detect this smaller impact in all the samples of students.

Two strands of analysis of the effects of the USBP will be pursued. The first is an analysis of the effects of the *availability* of the USBP. Should the USBP availability analysis, which benefits from the random assignment of schools (and/or students), not be able to detect impacts, then a fallback position for each design involves conducting an analysis of USBP participation.

a. Impacts Based on Analysis of USBP Availability

The availability analysis will estimate how the USBP influences outcomes among students at the USBP schools, relative to those at SBP schools. These outcomes include participation in the school breakfast program, breakfast consumption, dietary intake, and school outcomes such as student attendance, cognitive functioning and behavior, and academic achievement. Random assignment of schools implies that a simple comparison of mean outcomes between students in the USBP schools and those at SBP schools is an estimate of the effect of USBP availability on these outcomes. Regression analysis, however, will yield more precise estimates of this impact and will be used.

b. Impacts Based on an Analysis of Participation

Under the above plan for a randomized design, there is a fallback position. It is based on the fact that the evaluation can identify which students actually participate. This allows the evaluation

---

[8]Currently, approximately 30 percent of elementary students participate in the regular SBP. It is assumed that participation in the USBP will range between 50 and 60 percent, if the USBP is implemented in a traditional cafeteria setting. That is, it is expected that school breakfast participation will increase between 20 and 30 percentage points.

to directly model the effects of participation by comparing USBP participants and nonparticipants in the control group.

This line of analysis will estimate the impact of the USBP on those students who participate--that is, students who select school breakfasts in the USBP schools. The rationale for the SBP (and USBP) is that eating a nutritious breakfast will improve student outcomes, and that school breakfasts are, for some students, more nutritious than other breakfasts. To determine whether this rationale is correct, the participation analysis will directly estimate how eating school breakfasts influences student and health outcomes such as nutrient intake and academic achievement.

The participation analysis will focus on two basic sets of comparisons: (1) differences in outcomes between USBP participants and nonparticipants in the control group, and (2) differences in outcomes between USBP and SBP participants. The first set of comparisons looks at the effect of participating in the USBP itself, while the second set of comparisons looks at the relative effects of a regular school breakfast versus a breakfast provided under the USBP.

The methodology for the participation analysis will be to estimate a set of regression equations in which the dependent variables are the student outcomes of interest, and the independent variables include both school breakfast program participation and a set of control variables. This analysis lacks the advantage of a pure full experimental design, but it does allow considerable useful analysis. How well it will work depends on the degree of selection bias in students' participation analysis and the degree of the contractor's success in correcting for such bias. The random assignment design will help.

G. SUMMARY

A USBP is a policy option with potentially significant impacts on student well-being and performance. The evaluation of the USBP demonstration must be able to document these impacts.

Because of the wide range of potential impacts, the evaluation must be designed to capture both short-term intermediate effects on participation and nutrient intake and longer-term effects on cognitive, behavioral, and academic outcomes. The following chapters provide details on the approaches for the impact and implementation studies for evaluating the USBP demonstration.

PART 2:

THE IMPACT EVALUATION DESIGN

III. IMPACT STUDY RESEARCH QUESTIONS AND HYPOTHESES

Under a Universal-Free School Breakfast Program (USBP), school breakfasts would be free to all students, regardless of their family's ability to pay. Proponents of the USBP believe that, by making school breakfasts free to all students, the social stigma and financial barriers to participation will be removed, and, as a result, more children, especially those who might otherwise not eat breakfast or might eat nutritionally inadequate breakfasts, will consume breakfast at school. Since it is believed that school breakfasts are nutritious and improve student cognition and classroom behavior, it is expected that USBPs will favorably affect student academic achievement.

As discussed in Chapter II, except for breakfast program participation, the evidence on these relationships is at best suggestive, not definitive. Increased participation, while important, is only meaningful if it results in an increase in the nutritional quality of breakfasts consumed over what participating students would consume without the USBP. Earlier research, although suggestive of positive educational benefits, has not rigorously or conclusively demonstrated that universal-free school breakfasts or regular school breakfasts enhance student cognition and academic outcomes (Pollitt and Mathews 1998; and Briefel et al. 1999). The purpose of the USBP impact study is to determine whether the USBP pilot projects improve student dietary intake, attendance, cognition, and academic achievement.

This chapter describes the research objectives and questions of the impact study and the principal hypotheses to be tested.

A. IMPACT STUDY RESEARCH OBJECTIVES AND QUESTIONS

The objective of the impact study is to measure the effects of the *availability* of the USBP and of USBP *participation* on a broad range of school and dietary outcomes for elementary students.

Building on the conceptual model developed in Chapter II, nine primary sets of research questions guide the impact evaluation:

1. ***School Breakfast Participation.*** What is the impact of the availability of the USBP on students' participation in the school breakfast program? Does it increase participation on a given day? Does it increase usual participation?

2. ***Breakfast Consumption Patterns.*** How does the availability of the USBP affect elementary students' breakfast consumption? Are students for whom the USBP is available more likely to consume breakfast? Are they more likely to consume more than one breakfast on a given day?

3. ***Dietary Intake.*** How does both the availability of the USBP and participation in the program affect dietary intake at breakfast and over 24 hours? Does USBP availability and participation also affect usual intake at breakfast and daily intake?

4. ***Food Security.*** Does USBP availability and participation affect the student's household food security?

5. ***Attendance.*** How do USBP availability and participation affect student attendance? Are USBP participants absent from school and tardy less often? Do students participating in the USBP make fewer visits to the school nurse, and are they less likely to be absent from school due to illness?

6. ***Children's Health.*** What is the impact of USBP participation on children's health status, as reported by their parents? Is the intake of calories by USBP children higher than students under the SBP? Is there the potential for an increase in the prevalence of being obese?

7. ***Cognitive Functioning and Behavior.*** Does the consumption of USBP meals improve students' memory and positively affect their ability to perform tasks that require the retention of new information? Are USBP participants more attentive in class during the late morning; is time spent on-task greater? Are disciplinary incidents lower for USBP participants?

8. ***School Environment.*** How does the availability of the USBP affect the school's social climate?

9. ***Academic Achievement.*** Does USBP availability and participation enhance students' academic achievement? Do students with access to the USBP demonstrate greater gains in achievement on standardized tests than students who do not? Do USBP

participants experience greater gains in academic achievement than nonparticipants or SBP participants?

In addition to these main research questions, the impact study will address whether USBP availability and participation affect some subgroups of students more than others, focusing on key subgroups defined by such variables as family income (poor versus nonpoor), family composition, and nutritional risk.

Table III.1 reviews the data domains, measures, and unit of measurement for each research question listed above. All of the critical outcomes would be measured at the student level. In addition, for several outcomes, the evaluation will obtain both individual level and school level measures. The school-level measures would be student outcomes averaged over all the students in the school and averages by grade level. The data used to form the measures would come from (1) surveys of students (and their parents), including cognitive testing, in the treatment and control study schools; (2) school records data collection over time of sampled students' USBP and SBP participation, attendance, tardiness, disciplinary incidents, standardized achievement test scores, and other outcomes, including school-level information; and (3) surveys of sampled students' teachers.

B. PRINCIPAL HYPOTHESES THAT THE IMPACT ANALYSIS WILL ADDRESS

The impact study will examine the effects of the USBP on student outcomes. Prior research on USBPs provides some basis for identifying the impacts on student academic achievement and other outcomes most likely to be observed in the demonstration. Research on the regular SBP can also be used to inform expectations regarding the direction of effects. However, given the sometimes mixed and statistically insignificant findings, shortcomings in the research designs, and paucity of prior research, the formulation of specific, testable hypotheses concerning the impact of the USBP on some student outcomes must rely, in part, on a priori reasoning.

TABLE III.1

OVERVIEW OF RESEARCH QUESTIONS, OUTCOME MEASURES, AND UNIT OF MEASUREMENT FOR THE USBP EVALUATION'S IMPACT STUDY

Research Questions	Outcomes/Potential Measures	Unit of Measurement: School	Unit of Measurement: Student
What is the impact of the availability of the USBP on students' participation in the school breakfast program?	Participation rates in the USBP		
	Participate on a given day	X	X
	Participate during a given week	X	X
	Usual participation	X	X
How do the availability of and participation in the USBP affect elementary students' breakfast consumption? Are students with USBP available more likely to consume breakfast?	Breakfast consumption patterns		
	Eat breakfast on a given day		X
	Source of breakfast on a given day		X
	Usual breakfast consumption		X
How do both the availability of the USBP and participation in the program affect dietary intake at breakfast and over 24 hours? Do USBP availability and participation also affect usual intake at breakfast and daily intake?	Dietary intake		
	Nutrient intake at breakfast		X
	Nutrients selected from USBP/SBP		X
	Nutrient intake over 24 hours		X
	Usual intake at breakfast		X
	Usual intake over 24 hours		X
Do USBP availability and participation affect students' household food security?	Food security		X
How do USBP availability and participation affect student attendance?	School attendance		
	Number of days absent from school	X	X
	Number of days late for school	X	X
	Number of visits to the school nurse	X	X
	Number of absences due to illness	X	X
Does the consumption of USBP meals improve students' memory and positively affect their ability to perform tasks that require the retention of new information? Are USBP participants more attentive in class during the late morning? Is time-on-task greater? Are disciplinary incidents lower for USBP participants?	Cognition and behavior		
	Memory		X
	Attention		X
	Hyperactivity		X
	Emotional behavior/anxiety		X
	Disciplinary incidents	X	X
What is the impact of USBP participation on children's health status, as reported by their parents? Is the intake of calories by USBP children excessive compared with students under the SBP, resulting in overweight and obesity problems?	Health		
	Child health (parent report)		X
	Height/weight status (BMI)		X
How does the availability of the USBP affect the school climate?	School climate		
	Index of school climate		X
	Index of school problems		X
Do USBP availability and participation enhance students' academic achievement?	Academic achievement		
	Standardized test scores	X	X

BMI = Body Mass Index.

Reflecting uncertainty about the direction of several of the effects (that is, that the impacts may be either positive or negative), many of the hypotheses to be tested in the USBP evaluation will be of the generic form of a null and alternative hypothesis:

H_O: The change from the regular SBP to a USBP has no effect on student outcome "A."

H_A: The change from the regular SBP to a USBP changes student outcome "A."

Two tailed hypothesis tests would be used to test null and alternative hypotheses of this form. An impact estimate that is very much larger than zero, as well as one that is very much smaller than zero, would be evidence against the null hypothesis of no effect.

For some outcomes, however, it is reasonably certain that the impact of the USBP will be in one direction or the other. For example, on the basis of economic theory and prior research, it is expected that the switch from the regular SBP to the USBP will increase student participation in the school breakfast program, or that tardiness will decrease. In cases where it is reasonably certain that the direction of the effects will be in one direction or another, the hypotheses to be tested would be of the generic form:

H_O: The change from the regular SBP to a USBP has no effect on student outcome "A."

H_A: The change from the regular SBP to a USBP increases (decreases) student outcome "A."

The evaluation would use a one-tailed hypothesis test to assess null and alternative hypotheses of this form. In this case, an impact estimate that is very much larger (smaller) than zero would represent evidence against the null hypothesis of no effect.

The rest of this section discusses the hypotheses to be tested and expectations concerning the direction of the effects.

1. School Breakfast Program Participation

The impacts of the USBP on student outcomes, such as dietary intake and cognition, if they occur, will result from changes in students' school breakfast program participation. The evaluation will assess the impact of the availability of the USBP on participation in the school breakfast program. It is anticipated that school breakfast participation will increase in the planned USBP demonstration (see Table III.2). Prior research has shown that providing universal-free breakfasts increases participation in the school breakfast program (Wahlstrom et al. 1997; and Murphy et al. 1999a and 1999b). The actual magnitude of the increase in the USBP demonstration will depend on the setting in which breakfasts are provided: school breakfast participation is expected to increase more in schools in which a free breakfast is made a routine part of the school day, and, in particular, when it is linked to class-level activities, as opposed to when it is simply provided free to all students in, say, the school cafeteria, without additional interventions.

2. Breakfast Consumption Patterns

The evaluation will also assess the impact of the availability of the USBP and USBP participation on patterns of breakfast consumption. These breakfast patterns include whether students consume breakfast on a given day, their usual breakfast consumption, and the prevalence in which more than one breakfast is consumed during the day.

It is expected that, since school breakfasts are free, students with the USBP available will be more likely to consume breakfast on a given day; usual breakfast consumption also is expected to increase (see Table III.2). A possible unintended result of not charging students for school

TABLE III.2

EXPECTED EFFECTS OF THE USBP ON STUDENT OUTCOMES

Student Outcome	Expected Effect
Participation in the School Breakfast Program	
Participation on a given day	+
Usual participation	+
Breakfast Consumption Patterns	
Consume breakfast on a given day	+
Usual breakfast consumption	+
Consume more than one breakfast on a given day	+
Usually consume more than one breakfast	+
Dietary Intake	
Dietary intake at breakfast on a given day	+
Nutrients *wasted* from school breakfasts (plate waste)	-
24-hour dietary intake on a given day	+
Usual dietary intake at breakfast	+
Usual daily dietary intake	+
Food Security	+
Attendance	
Absences	-
Tardiness	-
School nurse visits	-
Absences due to illness	-
Cognitive Function and Behavior	
Memory	?
Attention/time-on-task	?
Hyperactivity	?
Emotional behavior/anxiety	?
Disciplinary incidents	?
Health	
Health status	?
Weight/height status	?
School Environment	+
Academic Achievement	
Scores on standardized achievement tests	?

NOTE: The signs in the table indicate the direction of the expected impacts of the USBP pilot projects, based on theory and prior research.

+ = increase
- = decrease
? = direction uncertain

breakfasts, however, is that it may cause students to consume more than one "breakfast." For example, they may consume breakfast at home, and then, since the breakfast is free, eat part or all of the breakfast available at school. Therefore, it is expected that students at USBP schools will be more likely than their student counterparts at regular SBP schools to consume more than one "breakfast" on a given day. A similar finding is anticipated when usual breakfast consumption of USBP students is considered.[1]

3. Dietary Intake

The USBP is expected to affect students' dietary intake in several different ways, including: dietary intakes at breakfast and over 24 hours (both on a given day and usual intake); the adequacy of nutrient and food group intakes; and the degree to which students meet dietary recommendations and standards. Also of interest are the nutrients selected-versus-consumed from school breakfasts. The difference between nutrients selected and consumed will provide an indication of "nutrients wasted" under the USBP, compared with the regular SBP.

Several studies have found that eating breakfast, including school breakfasts, has nutritional benefits (Devaney et al. 1995; and Devaney and Fraker 1989). Applying lessons learned from prior research on the regular SBP, it is expected that the USBP will result in an increase in the mean intake of nutrients both at breakfast and over 24 hours (see Table III.2). Similarly, usual breakfast intake and daily intake should increase.

It is important to note, however, that if increased dietary intake associated with the USBP leads to an intake of food energy that exceeds energy expenditure, then students may be consuming too

---

[1]The time of day when students consume breakfast could change under the USBP, depending on how the program is implemented. The timing of breakfast consumption will also be examined (see Chapter IV).

much food. For example, if new participants in the USBP simply add a school breakfast to what they normally would have eaten during the day (i.e., consume more than one breakfast), then food energy intake, both at breakfast and over 24 hours, will increase and may exceed energy requirements.

4. Food Security

The impact of the USBP on food security also will be assessed by the evaluation. It is expected that participation in the USBP will enhance the students' household food security status. First, students' food security should be directly enhanced because students who normally do not eat a school breakfast will now do so, and those who normally eat a school breakfast will do so more often. Second, household resources freed up from not having to spend money for students' breakfasts may now be used for other purposes, including purchasing food for other eating occasions and for other family members.

5. Attendance

Children miss school instruction if they are late to school or absent. Visits to the school nurse because of complaints of stomach aches and headaches that are due to children being hungry also disrupt children's learning. One of the most consistent findings of previous studies of school breakfast programs is that it positively impacts student attendance as measured by indicators of absences and tardiness (Pollitt and Mathews 1998). Evaluations of universal-free school breakfast programs have found significant reduction in absentee and tardiness rates (Meyers et al. 1989; Cook et al. 1996; and Murphy et al. 1998), as well as trips to the school nurse (Wahlstrom et al. 1997). The USBP evaluation will assess the impact on student attendance. Based on previous research, it

is anticipated that the USBP pilot projects will see a decrease in absenteeism, tardiness, and school nurse visits (see Table III.2).

6. Cognitive Functioning and Behavior

Prior studies suggest that *omitting* breakfast interferes with cognition, particularly those involving memory, and especially in at-risk children (Pollitt and Mathews 1998). No definitive conclusions, however, can be drawn from existing research on either the long-term effects of breakfast on cognition, nor on the effects of school breakfast programs. Some studies have found that children classified as "hungry" appear more likely to be anxious or depressed or to exhibit aggressive or disruptive behavior (Kleinman et al. 1998; and Murphy et al. 1998). In addition, three studies of the universal-free school breakfast programs have reported decreases in the number of disciplinary incidents at school (Wahlstrom et al. 1997; Murphy et al. 1998; and Abell Foundation 1998).

Proponents of universal-free breakfast programs argue that it will improve student cognition and classroom behavior. The belief is that students who skip breakfast or consume nutritionally inadequate breakfasts are unable to focus in class. It is possible, however, that a USBP may impair cognition and behavior: Students in USBP schools may overeat; they may eat breakfast at home and then eat some or all of the school breakfast that is available free at school. Such overconsumption could cause children to be lethargic and inattentive, and as a result, could adversely affect children's cognitive performance and classroom behavior.

Given the theoretical possibility that the effect on cognition and behavior could be positive or negative, and that prior research findings are inconclusive, the direction of the impact cannot be

determined a priori. The objective of the evaluation is to examine whether the impact of the USBP improves or impairs cognitive functioning and classroom behavior.

7. Children's Health Status

Research suggests that eating breakfast, including school breakfast, has nutritional benefits (Pollitt and Mathews 1998; and Briefel et al. 1999). Children who eat breakfast, either at home or at school, show improved intakes of dietary fiber, vitamins, and minerals. It is thought that breakfast consumption plays a role in improving overall nutritional status and health (Pollitt and Mathews 1998).

Again, when considering the potential impacts of the USBP, it is important to consider the potential for students to eat more than is recommended. Under this possible scenario, children may be more likely to be overweight. For example, in a study of a USBP program in a Philadelphia school, Murphy et al. (1999) documented a substantial prevalence of obesity in the sample.

Given this uncertainty, one cannot unequivocally predict the direction of the impact of the USBP on health. One of the goals of the evaluation is to determine the direction of this impact.

8. School Climate

A variety of underlying factors determine the climate of a school. These factors include students' attitude toward learning, their behavior in class, and the extent to which they feel safe and secure in school. Although school climate is a general term, it can be measured in surveys using questions that proxy for these underlying factors. For example, students or teachers can be asked about the extent to which "students in the school want to learn as much as possible," "disruptions by students get in the way of learning," or students feel safe in school."

Previous research suggests that the USBP has positive effects on school climate. For example, Wahlstrom et al. (1997) found that a large majority of teachers in five USBP pilot schools in Minnesota felt that the breakfast program "contributed" or "strongly contributed" to promoting students' learning readiness and social behavior. Murphy et al. (1999) found that more than half of the staff members surveyed in six USBP schools in Maryland had noticed improvements in students' attitudes or sense of community from the previous year. Given these findings, it is expected that the USBP will have a positive effect on school climate.

9. Academic Achievement

A major focus of the evaluation is to assess the impact of the USBP on children's academic achievement. Proponents of the USBP believe it will improve academic outcomes for more children. If students participating in the USBP improve their dietary intakes at breakfast, then it is possible to improve their cognition and nutritional health, and ultimately, increase learning potential and academic achievement. However, if participation in the USBP leads to overconsumption of food, the USBP may impair cognitive function and nutritional health, leading to a decrease in learning and achievement. And if, as a result of the USBP, children substitute nutritious breakfasts at school for ones previously consumed at home or from other sources, than the USBP would not affect student achievement.

Existing research on the effects of the USBP on student achievement is inconclusive. One study of the SBP has reported increases in some standardized test score percentiles over time. The findings are not definitive, however, for several methodological reasons. In particular, Meyers et al. (1989), a carefully conducted nonexperimental study, found positive effects on the CTBS composite score of about 10 percentage points of a standard deviation of the scores. Yet, the study failed to find significant effects on any of the CTBS subtests, suggesting that the effect on the battery total scale

could be due to chance. In addition, it appears the significance calculations may not account for the test that the sample is clustered in a small number of schools. Although the study controls for past test scores and some student characteristics, it most likely has not adequately controlled for the selection bias problem, and the sample inclusion criteria may have affected the study results. Finally, the intervention assessed by Meyers and his colleagues was substantially quite different from the USBP. The study evaluated the impact of introducing a SBP to low-income students who previously did not have access to a school breakfast, while the USBP offers a free breakfast, regardless of family income, to students who previously had access to the regular SBP. Other studies have found no effects of the USBP on student achievement. For example, Murphy et al. (1999), in Baltimore, found no greater increases in CTBS reading or math scores in the 31 schools that adopted a USBP program than in 15 similar comparison schools that did not adopt the USBP program.

Given the theoretical possibility that the effect on student achievement could be positive or negative, and that extant research is inconclusive, the direction of the impact of the USBP on academic outcomes cannot be predicted a priori. The objective of the evaluation will be to determine whether the USBP increases or decreases academic achievement.

IV. DATA DOMAINS AND OUTCOME MEASURES

This chapter describes the data domains that are important to the proposed evaluation of the effects of the Universal-Free School Breakfast Program (USBP) and the measures needed to implement the evaluation. The main decisions relate to defining and measuring outcomes, although the discussion includes the definition and measurement of key explanatory factors.

The USBP evaluation design will consider a broad range of programmatic outcomes. The legislation authorizing the USBP demonstration requires that the evaluation examine the program's impact on four outcomes in particular: (1) academic achievement, (2) tardiness and attendance, (3) dietary intake, and (4) breakfast program participation. In addition, the law requires a determination of the effect of the USBP demonstration on schools' paperwork. Examining changes in paperwork will be considered part of the implementation study (discussed in Part III of this design report).

In addition to the outcomes discussed above, the study will consider other possible effects of the USBP. For example, it may be easier and less expensive to detect impacts of the USBP on a child's attention and behavior on a given day or week than it would be to detect the program's effects on academic achievement over a longer period. In addition, the program's dietary impacts may translate into reduced illness and hunger. A key innovation of a USBP is its universal benefits. Therefore, a possible impact to measure is the reduction of stigma or stratification by income group within schools. The design options presented here provide guidelines for measuring these outcomes as well.

The chapter, organized in seven sections, recommends design features related to measurement of each type of outcome listed in the conceptual framework from Chapter II, such as breakfast

program participation outcomes, dietary outcomes, intermediate school outcomes, long-term school outcomes, health outcomes, and school climate outcomes. The final section of the chapter lays out the domains and measures for control variables that would be necessary, particularly under some design options proposed in this report, so as to better isolate USBP program impacts from confounding factors that also influence the outcomes of interest.

A. SCHOOL BREAKFAST PARTICIPATION OUTCOMES

Making school breakfast universally free is likely to affect individual students' participation in the breakfast program. The USBP could cause students who do not eat school breakfast to begin doing so, and it could cause those students who normally eat a school breakfast to do so more often. The net effect of these changes would be the impact on average regular participation. Estimating this impact would be important, both for predicting the costs of the program and for understanding the program's effects. Presumably, the impacts of the USBP on other outcomes result from increased consumption of school breakfast. It is hoped that the intervention will raise participation, which, in turn, will lead to improved diets, in terms of amount, quality, and timing of food intake. This diet improvement should lead to better attention, behavior, and health, and, ultimately, to improved academic achievement. In either case, measuring participation in the school's breakfast program (regular SBP or USBP) is key to the evaluation.

A comprehensive approach would measure participation (defined in a variety of ways) and the reasons for nonparticipation, to answer different research questions and test the robustness of the study's findings. Nevertheless, each data collection strategy may have different implications for how school breakfast participation is defined. The conceptual decisions to be made in defining participation are:

- What qualifies as a school breakfast
- Whether to measure usual participation, as well as participation on the target day (that is, the day of the dietary interview and cognitive tests)
- Whether to measure participation on the basis of food consumed or food selected

1. School Breakfast Definition

To measure participation, the USBP evaluation will have to decide what qualifies as a school breakfast. The most straightforward approach is to let the respondents themselves determine whether they ate a school breakfast by asking them a direct question, such as "Did you [or your child] eat school breakfast this morning?" Alternatively, the evaluation could define breakfast program participation based on the foods the school provides to the student in the morning. For example, the School Nutrition Dietary Assessment (SNDA-1) study defined school breakfast participants as those who obtained at least two items from the school cafeteria that contributed to the food pattern requirement. Regardless of how many servings are included in the definition, an additional issue for the USBP evaluation is how to treat foods that are provided through a la carte menus and vending machines.

The USBP evaluation will collect the appropriate data so that "participation" can be defined in a number of alternative ways. This will enable the evaluation to examine the robustness of impacts on participation, under alternative definitions of participation.

2. Target Day Participation and Usual Participation

Participation can be defined either in terms of a target day or of usual participation over a longer period such as a week, month, or school year. Measuring participation on the day of the interview will have less variability due to recall ability, but it may or may not accurately reflect usual

USBP/SBP participation patterns. Although the interviews will take place for different students on different days, the day-to-day variability for any given student remains a problem for same-day participation measures, particularly for analysis conducted at the individual-student level.

As discussed elsewhere in this report, the evaluation will include some analysis, with school breakfast participation as an explanatory variable. Therefore, which measure to use could also depend on whether the analysis focuses on the long- or short-term effects of a USBP. To study the effects of eating a school breakfast on tests of attention and behavior on a specific target day (short-term outcomes), the evaluation would need to measure participation on the day of testing. If the main interest is the effect of availability of the USBP on academic achievement growth during the school year (longer-term outcomes), then usual participation is more appropriate.

A hybrid approach between focusing on a single day and focusing on usual participation is to observe actual participation during a period of several days, such as a week. This potentially can be done by using observers (school cafeteria staff) at the cafeteria line, and it provides a measure that is not totally dependent on the experience of a single day but that also is not dependent on the reports of parents or students.

The evaluation will measure, or estimate, same-day, one-week, and usual participation in the school breakfast program. This would allow researchers to (1) remove the sources of variability, (2) answer a greater range of research questions, and (3) use the most appropriate variable for the time frame of interest in evaluating short- and long-term outcomes of school breakfast. For the target-day assessment, the evaluation will measure, in sufficient detail to allow multiple definitions, the amounts of each type of food taken from the school breakfast, and where the food was obtained.

As discussed in Chapter VI, usual USBP/SBP participation will be assessed through interviews with parents about students' frequency of school breakfast participation over the school year. This

would allow for estimates of the rate and frequency of school breakfast participation for the school-level analysis, as well as provide an estimate of usual participation for the individual-student-level analysis. We also recommend that observations of actual participation be taken over a one-week period.

3. Food Selected and Food Consumed

A student can select a complete school breakfast, then decide to waste food or share it with a classmate. The first student would be counted as a participant, while the classmate would not. Thus, one could define an alternative participation measure in terms of *consuming* certain amounts or patterns of food, rather than *selecting* certain foods. The dietary intake interview will be sufficiently detailed to answer questions about foods selected and foods consumed at breakfast for assessing students' dietary intake outcomes, while the school breakfast participation measure includes only foods selected on the target day.

B. DIETARY OUTCOMES

A major outcome of a school breakfast program like USBP is its effect on diet. Dietary outcomes of interest include whether the student ate breakfast, mean intakes of nutrients at breakfast and over 24 hours, adequacy of nutrient and food group intakes, and the degree to which students meet the dietary recommendations contained in the *Dietary Guidelines for Americans* and the USDA Food Guide Pyramid. Raw food intake data must be processed so that it can be expressed in terms of an outcome that is meaningful for policy--typically a measure of whether the intake is part of an adequate diet. Therefore, both qualitative and quantitative aspects of diet are important for this evaluation.

1. Breakfast Consumption

Devaney and Stuart (1998) showed that estimates of the effects of the SBP on the probability of eating breakfast using SNDA data are sensitive to how breakfast is defined. This may be true of the effects of the USBP as well. Breakfast can be defined in a variety of ways. Devaney and Stuart (1998) examined three major alternatives: (1) consumption of any food or beverage from waking until 45 minutes after the start of school, (2) intake of food energy of at least 10 percent of the RDA over the same period, and (3) consumption of food from at least two of five food groups and breakfast intake of food energy of at least 10 percent of the RDA.[1]

The evaluation will assess mean food and nutrient intakes for breakfasts consumed at home, at school, and for a specified time period of the day. In this manner, the 24-hour dietary recall provides flexibility in aggregating and disaggregating foods reported for various sources of breakfast and in assessing the contribution of home, school, and total breakfasts to the total intake over the course of the day. This will allow for comparisons between USBP/SBP participants and nonparticipants along a number of dietary parameters.

2. Dietary Intake

For the availability of a USBP to have any effects on nutritional status or school performance, the program must change what students eat. Thus, measurement of students' diets is key to the evaluation. Dietary intake over the course of the day is another outcome of central interest to the USBP evaluation. This includes evaluating breakfast consumption, as well as total daily dietary intake. The methodology that will be used in the USBP evaluation for estimating quantitative and qualitative intake is 24-hour-recall dietary intake interviews. Such interviews have been used to

---

[1]The five food groups are (1) milk and milk products, (2) meat and meat equivalents, (3) grain products, (4) fruits and fruit juices, and (5) vegetables and vegetable juices.

collect data on national samples of schoolchildren in the SNDA and the earlier National Evaluation of the School Nutrition Programs (NESNP-1), as well as in the major periodic national nutrition surveys--the Continuing Survey of Food Intakes by Individuals (CSFII) and the National Health and Nutrition Examination Survey (NHANES). The following subsections discuss conceptual issues related to measurement of dietary intake as an outcome variable. Operational issues related to dietary data collection are described in Chapter VI.

The following are relevant for assessing and interpreting dietary intakes for USBP/SBP participants and nonparticipants:

- ***Nutrient Standards.*** The adequacy of 24-hour dietary intakes of food energy, vitamins, and minerals has traditionally been assessed using the Recommended Dietary Allowances (RDAs) (National Research Council 1989).[2] Mean intakes typically are reported relative to the age- and gender-specific RDA. Currently, the RDAs are being replaced by the Dietary Reference Intakes (DRIs), which provide a broader set of standards for alternative analytic uses.[3] DRIs are not now available for all nutrients, but they will be soon. To the extent possible, it will be preferable to use DRIs in this study, since the DRIs represent the most current scientific knowledge concerning nutritional adequacy and recommended intake levels. The RDAs, however, should be used for nutrients for which DRIs are not available, and they may remain useful for comparisons with previous studies.

- ***Dietary Guidelines.*** Other important standards for assessing dietary intake are provided by the *Dietary Guidelines for Americans* (1995), which include recommendations for a healthy diet that apply to persons age two and above, such as eating a variety of foods, increasing fruit and vegetable intakes, and limiting intakes of total fat and saturated fat. These guidelines, which will be updated in early spring 2000, in time for the analysis in the proposed evaluation, have been utilized in the SNDA study and other studies. Additional recommendations of the National Research Council's *Diet and Health* (1989) could be used for those recommendations for which the *Dietary Guidelines* do not provide quantitative standards. The same cutoffs will be considered for use in this study, as well as those suggested by the new *Dietary Guidelines*. Again, it would be

---

[2]For breakfast intakes, USDA traditionally has set the goal that breakfast should meet 25 percent of the RDA.

[3]For example, Glazerman and Devaney (1998) illustrate how one DRI, the new Estimated Average Requirement (EAR), is superior to the RDA as a tool for measuring the prevalence of inadequate protein intake in a population.

possible to compare mean intakes to the standards or, if more than one day of intake is available, to assess the percentage of the population with "usual" intakes above or below the recommended cutoffs.

- ***Food-Based Standards.*** Another approach to meeting the *Dietary Guidelines* has been to recommend specific numbers of servings of foods from each of the major food groups. The USDA Food Guide Pyramid provides one way of grouping foods and gives recommended numbers of servings for foods in each group. It is difficult to estimate correctly the distribution of usual food intake in a population. Thus, it may not be feasible to estimate the percentage of students who usually consume less than the recommended number of servings of a food group. However, it is possible to compare the mean intake of specific food groups to recommended intake levels in the population.

In assessing the effects of the USBP on nutritional status as it relates to long-term outcomes, such as health and academic achievement growth, effects on *usual* dietary intake would be of primary interest. For example, high or low intakes on a particular day do not greatly affect a child's health, but high or low intakes over a sustained period of time can do so. However, as an intermediate outcome in assessing certain short-term effects of the USBP, such as effects on morning attention and behavior, it may be more important to assess intakes on a target day.

There is considerable variation in what individuals eat from day to day. Because of this, the distribution of intake on a given day is much broader than the distribution of usual intake. An important consequence is that collecting one day of dietary intake data for a sample provides an unbiased estimate of the mean intake of a nutrient for the population represented by the sample, but it does not provide accurate estimates of the distribution of intake of that nutrient or of features of the distribution such as the percentage of children below a cutoff value. Thus, a key conceptual issue for the design work is whether measurement of mean intakes is adequate or whether it is necessary to measure the percentage of the population who fall below particular cutoffs, in which case it is necessary to estimate the distribution of usual intake.

In practice, usual dietary intake cannot be observed, but the more days of dietary intake data available, the better it can be estimated. However, collecting dietary intake data on multiple days is costly and burdensome for the respondents. In recent years, researchers at Iowa State University have shown that it is possible to obtain good estimates of the distribution of usual intake for a nutrient with collection of as little as two days of dietary intake data (Nusser et al. 1996; and Carriquiry 1998). Unlike previous methods (National Research Council 1986), the approach of Nusser et al. allows for the fact that the distribution of intake for many nutrients is not normal (Gaussian) and is highly skewed.[4] The evaluation will collect two days of intake data for a random subsample of the larger evaluation sample, rather than for the full sample. Collection of two or three days of intake data for the full sample would make such estimates more precise but would be more costly, so is not recommended.

For target- or same-day estimates of dietary intake, information will be collected on both the foods selected for school breakfast and those consumed for school breakfast. In this manner, it will be possible to compare foods selected with food consumed, to evaluate (1) the potential additional dietary benefits or impact if the complete breakfast selected had been consumed, and (2) the degree of wastage of foods in the school breakfast program. The USBP/SBP nutrients wasted will be calculated by subtracting the nutrients for foods consumed for school breakfast from the nutrients for foods selected for school breakfast.

In summary, the evaluation will measure or estimate both same-day and usual intake to allow for estimates of means, as well as distributions of nutrient intakes so that dietary adequacy can be evaluated. When estimating usual intake, it is desirable to collect two days of intake data from the entire sample. Because resources are limited, the evaluation will collect the second day of intake

---

[4]Iowa State researchers also developed software for making these estimates, which MPR staff have used successfully in other research for FNS.

information from a random subset of the population, approximately 10 to 15 percent. The evaluation would then calculate, analyze, and report findings for food and nutrient intake using benchmarks such as the RDAs, the DRIs (subject to their availability), the *Dietary Guidelines*, and the Food Guide Pyramid.

3. Food Security

Complementing the analysis of breakfast participation and dietary intake outcomes, the evaluation will study improving food security. Simple hunger questions ("Did you feel hungry before breakfast today?" and "Did you feel hungry before it was lunchtime?) have been linked to breakfast consumption in a previous study (Wyon et al. 1997).

Using measures of food security, nutrition researchers have operationalized similar concepts at the household level. This will be useful for the USBP evaluation, since the availability of safe, adequate food in the household affects children's dietary intake and well-being. The 18-item food security measurement tool developed by FNS and National Center for Health Statistics in conjunction with researchers and experts in the field has been used successfully in the Current Population Survey (CPS), beginning in 1995 (Food and Nutrition Service 1997). The instrument allows for the categorization of households, and individuals within households, into one of four categories:

1. Food secure
2. Food insecure without hunger
3. Food insecure with moderate hunger
4. Food insecure with severe hunger

Because survey costs and interview time are of concern, the evaluation considered using the abbreviated, 6-item form, derived from the 18-item food security scale. It was rejected, however, for two reasons. First, using the 1995 CPS data, the short form had a lower sensitivity (86 percent) compared to its specificity (99 percent) for identifying overall food insecurity in households with children (Blumberg et al. 1999). Second, the short form does not distinguish between the third and fourth categories of severity of hunger (moderate or severe).

Therefore, the 18-item standardized food security instrument will be included as a measure of food insecurity and risk of hunger. This instrument will survey parents. While a child's self-report or an adapted version of the CPS instrument could also be considered an option for assessing food insecurity at the student level, such instruments have not been validated and therefore are not part of the preferred design.

C. INTERMEDIATE SCHOOL OUTCOMES

While the offer of a school breakfast to all students in the school may raise participation and improve the diets of students, the USBP demonstration evaluation also aims to measure whether these changes result in greater student learning. As is discussed in the conceptual framework of Chapter II, the mechanisms by which this can happen include the following:

- The program induces children to miss fewer days of school and to arrive promptly at school more often, thus exposing them to more time in class.
- The program's effect on diet makes students less disruptive and thus improves the quality of school time for them, as well as for the whole student body.
- The program's effect on diet makes students more alert and attentive and raises their cognitive functioning during the school day.

Because it may be difficult to observe directly the effects of USBP on achievement, the evaluation will also focus on those variables that mediate the relationship between school breakfast and academic growth. If the program has positive impacts on these intermediate outcomes, the ultimate effect of USBP on long-term learning can be further assessed using findings from the literature.

1. Attendance and Tardiness

The proposed evaluation design would examine the impacts of the USBP on student attendance and tardiness. By improving nutrition, eating a school breakfast should lead to better overall health and to a decrease in absences due to illness. The USBP may also encourage on-time arrival in order to participate in the program, thus reducing tardiness. An increase in the number of days students are present at school, and a decrease in disruptions to the educational program caused by tardiness, afford greater opportunity for learning.

Defining absence and tardiness is relatively straightforward. Schools routinely report some measure of average daily attendance to district and state education authorities; thus, obtaining these data at the school level should be relatively easy. Parental permission would be required for obtaining the data at the individual level. School districts may differ on whether they distinguish types of absences, such as unexcused absences or absences due to illness, but they are likely to have uniform reporting within districts. The most appropriate attendance measure would count days present in the school year, whether excused or not. The interest lies in how the program affects students' time in school, for whatever reason. An exception might occur if a school's record-keeping system made literal use of attendance records, without accounting for days where the educational experience of an "absent day," such as a field trip, is equivalent to a day of school. It is likely that excused absences are dealt with differently across schools or districts.

Tardiness is less routinely recorded in computerized school records, but most elementary schools keep a record of individual student tardiness (recorded by the homeroom or first-period teacher) and reported on the student performance report. Data on tardiness not available from centralized school records could be incorporated into a data collection instrument for the parent, teacher, or school. The definition of "tardiness" would have to be standardized. A reasonable standard definition might be: the student is late enough to miss class instruction or disrupt other students during instruction. Thus, arriving too late to eat breakfast but early enough to begin the school day should not be counted as a tardy day by the evaluation.

The evaluation will include measures of absence and tardiness based on student-level data, obtained from both student records and a survey-based measure. The evaluation will measure these outcomes in a way that allows researchers to interpret them as measures of productive time spent in school.

2. Classroom Behavior and Disciplinary Incidents

Student behavior in the classroom is another outcome of interest. Behavior is important, not only for the individual child's own learning, but for the overall classroom environment. Improving the behavior of one troublesome child in a classroom could benefit the teacher, everyone else in the class, and other students taught by that teacher by removing disruptions and raising the quality of the learning environment. A more long-term effect of improving students' classroom behavior might be to attract and retain good teachers. Among the aspects of behavior that might be important to learning and sensitive to nutrition are attention, aggression, impulsivity, and hyperactivity.

Potential instruments considered, listed in Table IV.1, include the Continuous Performance Test (CPT), the Child Behavior Checklist (CBCL), the Pediatric Symptoms Checklist (PSC), the Connors Teacher Rating Scale (CTRS), the Learning Behavior Scale (LBS), and the Mock Report Card.

TABLE IV.1

INSTRUMENTS FOR MEASURING SHORT-TERM BEHAVIORAL OUTCOMES

Instrument	Domain	Respondent	Comments
Continuous Performance Test (CPT)	Attention and Impulsivity	Child	May require expensive equipment
Child Behavior Checklist (CBCL)	Behavioral problems	Parent	Takes 30-40 minutes
Pediatric Symptoms Checklist (PSC)	Behavioral problems	Parent	Takes 5 minutes
Connors Teacher Rating Scale (CTRS)	Hyperactivity (subscale)	Teacher	Takes 10-15 minutes
Learning Behavior Scale (LBS)	Attention and Persistence (subscale)	Teacher	
Mock Report Card (teacher rating)	Interpersonal Classroom Skills, and Classroom Work Habits (subscale)	Teacher	

These instruments vary in the specific domain of behavior captured, the respondent surveyed (child, teacher, or parent), and a variety of other factors, including their acceptance in the research community and the feasibility of their implementation in an evaluation like the one proposed here.

The CPT is a direct measure of a child's attention and ability to inhibit his or her own behavior. This would be desirable to measure, since it would capture the same-day effects of school breakfast, but it is potentially costly and burdensome. It would normally be administered individually with a computer. Two parent survey instruments, the CBCL and the PSC, have been used in school nutrition studies. The PSC is considerably shorter and thus more likely to be feasible for the study currently proposed. The Mock Report Card is not a formal instrument, but rather a concept that could be adapted and included in the teacher survey for the USBP evaluation. It would seek to measure student behavior through standardized categories, such as student effort and conduct, that are often included in student report cards. While an intriguing approach, it is largely untested in research applications and is therefore not the best choice for the current study.

The CTRS is a commonly used tool for assessing children's behavior specifically in the classroom. Used for both clinical and research purposes, a recently revised CTRS (CTRS-R) has been introduced with various subscales measuring inattention-overactivity (IO) and aggression (A).

The CTRS has the advantages of being widely known and comparable to previous studies of children's behavior. Some concerns often raised about the CTRS are the test-retest correlations and internal consistency for the revised scale and potential teacher bias by student's race and gender (Epstein 1999; and Miller et al. 1999). Gender and racial bias are of concern for interpreting measurement levels, but for this evaluation, such bias may be less important in making impact estimates, since, on average, they would be based on populations of similar race and gender

composition. There is no a priori reason to expect teacher bias to be stronger in a treatment or a control school. The concerns about reliability are somewhat more pressing for this evaluation.

USBP availability could have effects on measurable student behavior patterns throughout the year. In addition to the more clinical definitions of behavior described above, the evaluation could measure the rate of disciplinary incidents over a longer period. Disciplinary incidents are more influenced by school policies and the specific relationship between school staff and the child than they are by separate clinical measures. On the other hand, quantitative measures of incidents are easier to construct. Possible measures include number of office visits, detention days, classroom "time-outs," suspensions, or expulsions per month. The availability of data on disciplinary incidents depends on whether schools or districts selected for the demonstration keep records that are comparable within the school district. Comparability across school district would be desirable, but treatment-control impacts in terms of standardized units can be aggregated even if the levels cannot.

At least one instrument designed to measure student behavior directly, and one to measure disciplinary incidents, will be included in the design. It is recommended that the evaluation use the CTRS-R. The CTRS-R is based on teacher ratings. Since the types of disciplinary infractions most common in elementary populations may not be systematically recorded by the school, the evaluation may need to include items on the teacher survey to measure this domain. If the districts selected for the demonstration have consistent student discipline policies and comprehensive record keeping, then the evaluation should also use administrative data in this component of the research.

3. Cognitive Functioning

Another direct contributor to a child's ability to learn is his or her cognitive functioning. Developmental psychologists have produced a variety of assessment instruments covering many

domains of possible interest to the study. Available instruments, listed in Table IV.2, cover such subdomains as:

- Visual perception
- Verbal memory
- Verbal fluency
- Time-on-task

It will be desirable to relate these outcomes to same-day participation, if possible. It also will be important to choose the instruments that provide valid, reliable measures at the lowest possible cost of materials and interviewer training.

The experimental literature on the effects of breakfast (Pollit 1995; and Vaisman et al. 1996) has shown that tests of verbal memory, such as the Wechsler Memory Scale and the Benton Visual Retention Test (BVRT), are sensitive to breakfast consumption. In addition, recent evidence on small samples suggests that breakfast may be related to time on task, the fraction of time a student can concentrate on a given school activity (Cueto et al., in press). The study by Cueto and colleagues used videotape and analyzed the fraction of class time that students spent looking at the blackboard. Since the costs of videotaping students in the USBP evaluation would be prohibitive, an alternative would be to include an item on a teacher survey that asks the same thing. For example, "For what fraction of a today's class period did [child's name] pay attention to the task at hand?" The choice of instruments should be guided in part by feasibility, given the other components of the study.

Ideally, one would like to include in the child testing instrumentation measuring each of the four subdomains highlighted in the bullet list above. Candidates would include the Matching Familiar Figures Test, the Wechsler Memory Scale, and a subtest of the Clinical Evaluation of Language

TABLE IV.2

INSTRUMENTS FOR MEASURING SHORT-TERM COGNITIVE AND EMOTIONAL OUTCOMES

Instrument	Domain	Respondent	Comments
Cognitive Functioning			
Matching Familiar Figures Test (MFFT)	Perception	Child	
Stimulus Discrimination	Perception	Child	Takes 10-15 minutes; see Pollitt, Cueto and Jacoby 1998
Rey Auditory-Verbal Learning Test	Learning and Memory	Child	Used by Vaisman et al. (1996)
Wechsler Memory Scale	Verbal Memory	Child	Used by Vaisman et al. (1996)
Benton Visual Retention Test (BVRT)	Visual Memory	Child	Takes 10-15 minutes
Time on Task	Concentration	Teacher	New question developed for USBP evaluation.
Kagan Test	Verbal Fluency	Child	Used by Simeon and Grantham-McGreggor (1989)
Emotional Symptoms			
Revised Children's Manifest Anxiety Scale (RCMAS)	Anxiety	Child	
Children's Depression Inventory (CDI)	Depression	Child	

[a]Time estimate assumes using only the verbal memory scale component of the longer full test.

Functioning. However, these tests take approximately 10 to 15 minutes. Because of the young ages of the children to be tested, the need to conduct a short dietary recall, and possible limits to the willingness of teachers and school administrators to provide access to the students, it is not feasible to include all the subdomains. Therefore, one cognitive test will be administered--the Wechsler Memory Scale. This test is recommended because it has performed well in research settings and is relatively straightforward to administer.[5] The evaluation also will include items on the teacher survey to ask about each sampled child's usual attention in class. This question should be worded in a way to elicit a percentage of the class time during which the child usually pays attention to the task at hand.[6]

4. Emotional Effects

Another intermediary variable that could be affected by the USBP involves emotional issues for the students. While these outcomes may not be as closely related to achievement as test scores, they still are of interest and have been shown to be sensitive to children eating breakfast. Key areas of interest, for instance, are potential effects on children's feelings of depression and feelings of anxiety. It would be desirable, therefore, to include a module to obtain information on this issue with the student survey.

Instruments for measuring anxiety and depression are listed in the second part of Table IV.2. Researchers have found effects of school interventions on student reports of emotional symptoms

---

[5]If statistical precision were not an issue, it would be tempting to undertake a split-sample design, administering each of the four recommended tests to subsamples of the overall sample. However, in light of the concerns about statistical precision raised in Chapter V, further subdividing the sample would not be desirable.

[6]If sufficient time and resources are available, it would be desirable to conduct a pilot study based on administering all four tests in common settings, in order to test their feasibility and perhaps develop preliminary hypotheses as to which best meet the needs of the current study.

using the Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds and Richman 1985) and the Children's Depression Inventory (CDI; Kovacs 1985).

The evaluation will include one module on emotional well-being. Either of those listed in Table IV.2 would appear to be suitable, and the choice should be driven largely by which area of emotional health (depression or anxiety) is believed to be most likely to be affected by school breakfasts. With no prior belief concerning which area is more important in this context, it is recommended that the evaluation use the RCMAS, because it is somewhat shorter and easier to administer.

D. LONG-TERM SCHOOL ACHIEVEMENT OUTCOMES

A major question for the USBP demonstration is whether the intervention contributes to student learning. The most direct way to answer this is through test scores, to measure the program's impact on academic achievement. While this is perhaps the most interesting policy question, a realistic assessment of the task reveals that it is extremely challenging. Obtaining valid, reliable measures of academic achievement is difficult. Even then, the variation in achievement test scores--due to student abilities, family influences, and the myriad differences in education policies and interventions across schools--may swamp any effects of USBP availability. Furthermore, the direct impacts related to eating a nutritious breakfast will be diluted in the USBP study because the evaluation cannot reliably identify ahead of time which students will already be receiving an adequate breakfast, through the regular SBP or home meals. Including all students means averaging the outcomes of those influenced by the program with outcomes for everyone else. To overcome this problem, particularly with regard to achievement, the evaluation design should make every effort to increase the precision of the test score impact estimates. That way, if the estimated program impacts are not statistically significant, the results will still be informative.

A key issue, and one that cannot be resolved until FNS learns which SFAs apply for the demonstration, is whether the evaluation contractor will need to conduct follow-up achievement tests. Testing children is costly; it also imposes a significant burden on children and teachers. The best-case scenario for the evaluation is if school districts selected for the study already had in place an adequate system of achievement testing that satisfied the needs of the USBP impact analysis. In that case, obtaining achievement data would be merely a question of extracting records from school databases and, if necessary, securing parental permission for individual-level records. In some cases, historical data would be available to add precision to the analysis.

The preferred approach to measuring student achievement in the USBP evaluation is to use data on existing tests administered by participating school districts. It is assumed that all participating districts will be able to provide a measure of preimplementation achievement for sampled students. If districts can also provide follow-up data on sampled students, then the evaluation would not have to conduct its own achievement test. This would free resources to increase the school and student analysis samples, which would increase the precision of the impact estimates.

Much of the discussion in the rest of this section applies to the case in which the evaluation would need to conduct a single round of follow-up testing on the students in the longitudinal sample. In that case, the following conceptual issues must be addressed:

- What domains should be included
- Who should administer the test
- Who should be tested and how often
- What instrument should be used

1. Achievement Domains

Most achievement testing in elementary grades covers the same basic subjects: core skills and content knowledge areas such as reading, writing, spelling, and arithmetic. Some tests stress thinking skills and problem solving, while others place greater emphasis on math computation and language mechanics. Many widely used tests try to measure a combination of these subjects through a variety of test item formats, as discussed below. A typical test battery may have subtests that focus on different areas. For example, the California Achievement Test (CAT) for elementary students has separate subtests for math computation, math concepts, reading, and vocabulary. Ideally, for parents and educators, it would be important to measure every domain that is a valued part of the student's education, in order to form a complete picture and correct any deficiencies at the individual level.

For the USBP evaluation, however, it will be necessary to select a narrower domain of test items as a way to lower costs and ease the burden of a full battery of tests. For example, with the CAT, one could use only the mathematics concepts and reading subtests. As long as the tests used are good proxies for the subject areas not included in the analysis, this would be a cost-effective evaluation design.

The evaluation should attempt to select a relatively narrow set of test domains as a way to lower costs and ease the burden of collecting achievement data from students. To do this, but also to include the most important domains of achievement, tests administered by the evaluation contractor should include a verbal component and a mathematical component.

2. Achievement Test Instruments

When deciding how to measure achievement, a key question concerns what instrument to use. Here the focus is on choosing among standardized tests.[7]

If chosen carefully, standardized tests can be used effectively to make comparisons for evaluation of the USBP. The types of standardized tests commonly in use are norm-referenced tests (NRTs) and criterion-referenced tests (CRTs). Within these types are many potential instruments. Instruments from each type can consist of selected response (for example, multiple-choice) items or constructed-response (essay) items.[8]

A norm-referenced test is one that is designed so scores can be compared to a reference group of students--typically, the nation as a whole. Publishers of NRTs traditionally administer the test to a large, nationally representative sample of students of a given grade level, called the "norm group" because it represents the norm population. This is customarily done every 5 to 10 years. A score that ranks in the 85th percentile, for example, implies that the student would perform better than 85 percent of all people in the norm population.

Many states and school districts administer CRTs instead of NRTs. A criterion-referenced test is designed to compare a given score with a substantive content goal. Thus, for example, a score

---

[7]Another logical possibility for measuring student achievement is school grades. However, while it is likely that all teachers in the demonstration schools would assign some kind of grades to their students on a regular basis, it is unlikely that such grading systems are comparable across school districts, schools, or even classrooms. Comparisons over time may also prove difficult, particularly if school staff and policies are changing. Furthermore, classroom grading at the elementary level often paints with a broad brush, assigning students to three or four qualitative categories whose exact meaning must be interpreted on a case-by-case basis. Gathering data on all classroom grades and recording them in a common format for comparative analysis is a formidable task. Therefore the evaluation will not select grades as an outcome. The focus of the rest of this discussion is therefore on standardized tests.

[8]In popular debate over student testing, the term "test" is sometimes reserved for selected-response instruments, while "assessment" is used to denote constructed-response instruments; however, the terms are used interchangeably here.

might indicate whether the student has "mastered long division." CRTs are more popular with educators and parents than are NRTs because CRTs have a more immediate interpretation. CRTs, however, are less useful for the evaluation because the impact analysis is interested in differences in groups who did and did not have the universal availability of free school breakfast.[9] Absolute levels are less important in this context. Furthermore, focusing on CRTs would complicate efforts to pool data or make comparisons across school districts. In particular, CRTs can be used if all test takers are taking the same curriculum and the same test, but with multiple districts that presumably are drawn from multiple states, this is unlikely to be the case. Therefore, NRTs are preferable to CRTs for the current evaluation.

A standardized test does not necessarily have to consist of multiple-choice items. Item formats such as multiple choice and true/false are sometimes called selected response formats. Increasingly, however, test publishers are developing tests that also ask students to construct their responses instead of choosing a given response. For example, the test taker may have to respond to an open-ended question, justify a calculation, write an essay, or physically demonstrate a skill. The choice of test-item formats is often a subject of bitter controversy. Most are not inherently more or less suited to the purposes of studying the USBP.

For the USBP evaluation, contractor-administered tests should be selected according to the following criteria:[10]

---

[9]One type of CRT is a minimum competency test. Minimum competency tests are designed to give a score that is above or below some criterion threshold, without distinguishing among other performance levels. Since many students in the demonstration may not be near the cutoff point, minimum competency tests are less useful for the evaluation.

[10]Here, unbiasedness is not listed as a critical criterion because, under certain evaluation designs, the bias would drop out of the analysis if it operates equally on treatment and control groups.

- ***Is the test valid?*** Constructed-response items tend to have the advantage in this area, particularly in content validity (whether the items measure something important). Selected-response items sometimes have high predictive validity, meaning they are good predictors of future performance.

- ***Is the test reliable?*** In other words, would multiple administrations of the test give about the same result, and would the same test scored by different people give the same score? Closed-end questions tend to have higher reliability, primarily because they are machine-scored and not subject to the inter-rater differences that plague essay and performance scoring.

- ***Is the length appropriate?*** Each item on a multiple-choice test takes less time to administer and score, although more items may be needed to generate a valid estimate of achievement. These considerations are important because too long a test could be costly and burdensome. A subset of test items could give a good prediction of what the score on the longer test would have been, but a short test could also be too imprecise and discriminate over too narrow a range of achievement levels to be useful for the evaluation. Test publishers often sell modules that can be combined to create a test of the desired length.

- ***Is the test discriminating?*** Any given test item typically will discriminate between achievement levels specific to the difficulty of the item. For example, a difficult test question is useful for separating a high achiever from the rest of the population, but not for determining whether someone has or has not met a minimum competency standard. Therefore, many items, some difficult and some less challenging, would be needed to be discriminating among low- and high-achieving students. Failing to include a wide range of difficulty in the items could result in a "floor" or a "ceiling effect." A test with floor effects is one where low and very low achievement levels are indistinguishable from each other because the items are too difficult for the test takers. A ceiling effect is one that fails to distinguish between high and very high performance because there are not enough challenging items for the test takers.

Based on these criteria, any of the following testing packages would serve the interests of the study well:

- The Terra Nova standardized tests
- The Iowa Test of Basic Skills
- The Comprehensive Test of Basic Skills

All three of these commercially available testing packages offer subsets of tests that can be administered in the elementary grades in a period of one to two hours. All are carefully normed and widely used, and all have face validity within the evaluation community.

If the evaluation contractor must administer an achievement test in the USBP evaluation, then it is recommended that the Iowa Test be used. However, it may be prudent to postpone a final decision until it is determined what testing packages the school districts selected for the evaluation currently use. If one of the above tests is used predominantly in those districts, then it may be appropriate to choose that package for the follow-up test, in order to allow greater comparability across sites.

Another possibility to consider is "adaptive testing," an approach that conveniently addresses several of the above concerns. An adaptive test begins with a very large item bank that includes test questions designed to measure a single scale that goes from very low to very high achievement. The testing procedure often begins by giving each student a "locator test" to gain an initial estimate of the student's approximate achievement level. Then, based on the student's score on the locator, the student is given additional questions that are pegged to his or her estimated achievement level. If the student answers questions correctly, then harder questions are given. If the answer is incorrect, then slightly easier questions are given. This process continues until the student's achievement can be estimated within some preset margin of error. Often a precise estimate can be obtained quickly, using far fewer items than would be necessary if a single test had to be used for all children at all achievement levels. This reduces test-taking time, burden on students (because the questions, by design, present just enough challenge so the student does not lose hope or become bored), and floor or ceiling effects (because there is no constraint on the range of items included in the item bank).

Adaptive tests are often made feasible by using computer administration. That way, each item can be scored as it is answered and the test can be made "perfectly adaptive." A paper-and-pencil version gives all students a short locator form, which is collected and scored by computer in an adjacent room, while instructions are being read for the second part. The second part is then given out, but test takers receive different test booklets for the second part, depending on their initial scores. This type of administration is sometimes referred to as "mildly adaptive."

In summary, should the evaluation contractor need to conduct a follow-up achievement test, the USBP evaluation will rely only on NRTs. The item format is not critical, but the tests must meet prevailing standards of validity and reliability, and should be of "reasonable length" to minimize burden on test takers. Furthermore, the test instruments should be discriminating over a wide range of achievement levels, to rule out floor and ceiling effects and to allow for aggregation across different ages, grade levels, and stages of cognitive development. At this time, it is assumed the evaluation will use the Iowa Test of Basic Skills. Adaptive testing, while attractive, is probably not feasible, because of logistic issues.

3. Who Should Administer the Test?

If the evaluation contractor must administer a follow-up achievement test, then it must be decided who should administer the achievement tests. The answer depends partly on logistics and cost. One option is for the evaluators to administer a test to sample members; a hybrid option would be to rely on existing district tests but supplement them with resources from the evaluation.

Supplementing District Assessments. Even if districts participating in the USBP demonstration use NRTs, their assessment policies may have shortcomings that threaten the quality

of the data used for the evaluation. For example, the following problems could arise with school districts included in the study:

- They might not test frequently enough.
- They might not test at enough grade levels.
- They might not test enough subjects.
- They might not have high enough completion rates.
- They might not adhere to the published test administration guidelines.
- They might not score tests appropriately for the evaluation.

It is possible that each of these problems can be overcome through negotiation with the districts and the use of evaluation funds to assist districts in their compliance with study protocols. The evaluation could consider the following levers:

- Provide a fixed subsidy per sample member in the district
- Provide staff support for proctoring or scoring
- Provide summary reports on value added to achievement, by school
- Provide technical assistance with implementation of an accountability system or program evaluation
- Provide technical assistance with selection and interpretation of tests
- Provide extra funds for baseline testing of new students migrating into the system

If these measures are not sufficient to gain district cooperation and ensure a rigorous measurement protocol, then the evaluation contractor team would have to consider administering its own achievement test to sample members during the follow-up period. This option is considered next.

Evaluator-Administered Test. The USBP evaluation could achieve the highest comparability and data integrity by selecting, administering, and scoring a follow-up achievement test of its own choosing. A major drawback to this would be costs. In addition, if the contractor could not persuade schools to allot class time for testing, a problem in administering special tests for program evaluation purposes is the lack of incentives for study participants. Students and their families would need to be induced to appear at the testing site, which, even with sizable respondent payments, could have a high attrition rate.

The other problem is ensuring that students take the test seriously. Tests that are part of the districts' current assessment system are more likely to carry stakes that give test takers motivation to do their best. There is no guarantee that sample members who take a written test designed for the USBP evaluation will not provide bogus answers or haphazard guesses to test questions. While biases due to lack of student effort might also drop out of a treatment-control difference, they would nevertheless dilute the statistical power of the test.

4. When to Measure Achievement

For the USBP evaluation, academic achievement can be thought of as a long-term outcome, the result of learning over the entire school year, as opposed to an effect of school breakfast that might take place on the same day of a given food intake period. For that reason, it would be useful to have measures of student achievement at the beginning and end of each year. However, since school districts typically administer their own achievement tests in the spring, at the end of the school year, a reasonable compromise is to use spring-to-spring results to estimate such year-to-year changes, as discussed below.

As discussed in Chapter V, substantial gains in precision can be realized by using prior achievement (pretest) as a control in the analysis of posttest scores For that reason, a fall pretest

would be ideal because it would allow the researchers to characterize growth over an entire academic school year without losing precision due to changes over the summer before the USBP program year that are obviously not attributable to that program year. A fall pretest would also eliminate some of the problems resulting from student mobility that would translate into study attrition. The problem, however, would be that tests in fall and spring would be twice a year, instead of spring to spring, which would be once a year. Using the posttest from one year as a pretest for the next year is far more efficient, and probably outweighs the benefits of fall-to-spring testing. In addition, fall testing can provide the wrong incentives for teachers, who might realize that the test is a pretest, or to students, who would have no reason to perform well.

Measures of student achievement for at least two points in time will be obtained. The feasible schedule will be to collect data for the spring.

5. Grade Levels

Using pretest data requires that the evaluation have test scores for students adjacent at grade levels. Thus, if students are tested in grades 2 through 6, the evaluation could analyze results from students in grades 3 through 6, since second graders have no pretest. Under the scenario where the evaluation conducts its own follow-up test, it may be necessary to test students in only one or two grades. This would mean that the evaluation would not be able to draw conclusions about achievement impacts at other grade levels treated in the demonstration, but that power would be added to the analysis.

Focusing on one or two grade levels also avoids the difficult problem of aggregating across grade levels. A well-designed test would be "vertically equated," which loosely means that tests for each grade level measure an underlying growth process that can be expressed in the same metric for children of different ages and stages of development. Without this property, any combination of

program impacts from different grade levels would be biased in unknown ways toward one of the grade levels, purely as an artifact of the test.

There are some practical considerations when choosing grade levels to test. First, grade configurations across schools and districts may differ. It is estimated that of all public elementary schools in the United States with a first grade, about 89 percent go through at least grade 5 and about 44 percent go through at least grade 6.[11] Therefore, it seems safe to assume that all schools in the evaluation will likely have grades 1 through 5, but that many will not include grade 6 also.

The final decision about which grades to include in the analysis sample cannot be made until FNS finds out which SFAs apply for the demonstration, when it is learned what are the grade configurations of participating schools and the availability of district-administered achievement test data. If the tests used by participating districts are not vertically equated, then the evaluation may have to restrict the focus of the achievement analysis to a single cohort of students over a single year. Otherwise, the most likely grades to include would be grades 2 through 5 during demonstration implementation.

E. HEALTH OUTCOMES

The USBP evaluation will also examine the impacts of the program on improving the health of students. Increased participation in the USBP may lead to improvements in overall nutrition and, ultimately, to better health. There are several ways to measure health outcomes, including parent reports, nurse visits, and school days missed due to illness. Anthropometric data, such as height and weight, can also indicate students health. These data may be available from school or district

[11]These numbers are from calculations based on tables provided by the National Center for Education Statistics, Common Core of Data.

records. However, not all schools or school districts maintain the number of nurse visits, and many keep a tally of total absences without recording the cause (that is, for illness).

1. Self-Reported Health

A measure that uses parent reports of child health can be more standardized across the study despite its being a subjective measure; it will be included in the parent survey. The use of Likert scaled questions of this nature is a common approach to assessing child health, including national governmental surveys such as the National Ambulatory Medical Survey and individual child assessment forms like the Child Health Questionnaire (Landgraf et al. 1996). An Alameda County study has consistently found significant relationships between reported health status and breakfast consumption in adults (Bellow and Breslow 1972; and Kaplan 1986).

2. Number of Visits to the School Nurse

Schools that normally do not maintain the number of visits to the nurse may be willing to comply with an evaluation study request--particularly if they are asked to keep totals rather than visits by individual students. It may be easier and less burdensome to request school averages for nurse visits.

3. Absences Due to Illness

Student-level attendance records are readily available from school records, but parental and district permission will be required to release this information for individual students. School averages will also be obtained for attendance, since the information is part of the schools' reporting requirements to district and state authorities. Questions about the child's absences due to illness will be included in the parent survey.

4. Height and Weight

Another measure that captures a child's health is body weight. The USBP evaluation will track the effects of the program on weight gain, controlling for height. If taken by the study team, weight and height measurements could yield accurate and consistent data. On the other hand, this measure would represent an added, intrusive component to the data collection.

In summary, the evaluation will collect data on student health based on school administrative records and include "self-reported" health questions on the parent survey. It will also include measuring students' heights and weights during the follow-up interview, if such data are not already available from school records.

F. SCHOOL-WIDE OUTCOMES

In addition to influencing the diet, health, and learning of students individually, the USBP has the potential to change the school environment. Here we consider potential measures of change in the school climate and prevailing attitudes as a result of USBP.

1. School Climate

Programs that target disadvantaged students tend to segregate them, and thus allow a stigma to develop around receipt of services. A unique aspect of the universal-free school breakfast program is its universality, the detachment of free breakfast from any income requirements. Therefore, a possible effect of the program could be to reverse any such segregation and stigma. The proposed design supports a study of program effects on stratification and stigma, since it uses entire schools as units of assignment. Measuring the impacts of the program on these outcomes requires subjective assessments, but it is worth conducting these assessments. Although "school climate" is a general term, it can be measured in surveys using questions that proxy for these underlying factors. The

evaluation will use student interviews from USBP and control schools, to compare responses to questions about stigma. The instrument used in SNDA included short questions on children's perception of the breakfast program. School staff reports of differences in school climate is another way to measure this outcome.

2. Attitudes

Over the course of intervention, students, administrators, teachers, and parents will develop both attitudes toward the USBP and perceptions about its impact. Attitudes toward the USBP program will reveal both levels of satisfaction with the program operation and its perceived impact on the promotion of a positive learning environment.

In summary, the issue of stigma and stratification, as well as attitudes toward the program, will be addressed by the evaluation through the process evaluation by inserting questions in the student, parent, teacher and school administrator surveys.

G. PROPOSED CONTROL VARIABLES

An important part of the data collection and analysis will be to include measures of other explanatory variables that contribute to different outcomes. This information will not only help increase the precision of program impact estimates by reducing the variance in outcomes, but will strengthen the validity of any nonexperimental analyses by controlling for alternative causes of outcome differences that might otherwise be confounded by school breakfast participation. Some of these control variables pertain to the individual level, while others are collected at the school or SFA level. Individual-level statistics can always be aggregated to the school or district level. However, with a fixed number of schools--say, 144--there will be limited degrees of freedom for analysis; therefore, investigators may have to choose carefully among control variables at the school

level. District-level information can be collected in rich detail through discussions with the six SFAs, presumably as part of the process evaluation analysis.

The variables are grouped here by the outcomes they are hypothesized to affect. For each, the relevant measurement issues that would affect the study design are discussed.

1. Key Determinants/Predictors of Individual Cognitive Achievement

A vast literature on educational achievement, including estimates of the "education production function" (Hanushek 1986; Coleman et al. 1982; and Goldhaber and Brewer 1997), provides evidence on the determinants of academic outcomes. Not surprisingly, the most important of these predictors of achievement is prior academic achievement. To be useful for this study, pretest measures do not necessarily have to be in the same metric as the posttest score, or even in the same subject area, although it would be best if they were. The issues associated with measuring prior achievement are dealt with above.

The question then becomes: What are the determinants of achievement growth from one testing occasion to the next? These can be grouped into student and family background characteristics, peer group characteristics, and school characteristics. Most of the student and family background characteristics can be measured as part of a parent or child interview. Perhaps the most important of these, in terms of predictive power, would be the education level of the student's mother. The effect is somewhat stronger for the mother's education than the father's, due partly to stronger maternal influences on children. Since her education and that of the father are often highly correlated, measuring only the mother's education level is usually sufficient to capture this variable. Another reason for using the mother's education is that nearly all students will have a mother or female guardian in the home, whereas many will not have a father or male guardian. This question

of maternal education level could be asked of a parent who assists with a child's dietary intake interview.

Other family background measures that describe socioeconomic status would be helpful, including annual household income, parents' occupations, household composition, and language spoken. Whether a child lives with two parents, how many siblings the child has, and the parents' employment status--all could affect the amount of time and attention the child receives from adults in the home; adult time and attention influence children's development.

An important student characteristic that would influence test scores is disability status. Information on student disabilities tends to be recorded universally, since federal funding for compensatory education requires districts to establish an Individualized Education Plan (IEP) for any student with a disability. The challenge for the USBP evaluation will be gaining access to records indicating whether sampled students have an IEP, or gaining access to the IEP itself. The level of detail available to the USBP study will depend on district record keeping, as well as on the ability of the study to ensure confidentiality and secure cooperation from stakeholders.

Some demographic variables may also be useful for increasing the precision of the effect estimates. Normally, for example, race, ethnicity, and gender would be included in administrative or interview data. These characteristics would be useful control variables for all analyses, as well as for studying subgroup impacts.

2. Key Determinants of Attendance and Tardiness

In addition to the factors just listed, commuting distance or commute mode, as well as details about parents' employment, could be useful for reducing variance in the analysis of attendance outcomes. Attendance rates may depend somewhat on whether the student takes the bus, gets a ride, or walks to school, and how long it takes to get there. This information could come from an

interview with the student or parent. Attendance may also depend on whether someone is at home to make sure the child is ready for school or to provide child care for students who miss school. Therefore, information on parental employment and daily commuting patterns will be part of the parent interview.

3. Key Determinants of Individual Nutrient Intake and Health Outcomes

Given the main hypothesis that school breakfast affects dietary intake and health, it is useful to control for other food assistance programs. The parent interview, then, will ask about participation in the National School Lunch Program (NSLP) and the Food Stamp Program (FSP). Issues with measurement of school lunch participation are similar to those for measuring USBP participation. One definition could use the usual consumption of school lunch, which would require a survey question similar to one asked about usual breakfast participation. Another definition could use consumption of school lunch on the day of the intake interview; this could be part of that intake interview.

Participation in the FSP can be asked of parents in an interview. Following the CSFII question format, the items would ask whether the household has been certified to receive food stamps in the current month, previous three months, or ever. There are advantages to knowing the benefit amount as well.

Another important control variable for studying nutrition and health is body mass index, the ratio of weight to height. How carefully these need to be measured would depend on how they will be used in the analysis. One method is to use school nurse's records, although it is unlikely that schools would have such records available on all students in the study, and extracting that information could be more costly than direct measurement. Another method would be to take direct measurements during either the student interview in school or the intake interview in the home. This requires interviewers to have consistent instruments.

4. School-Level Variables

Finally, it will be useful to measure selected characteristics of each school. The following list of characteristics includes many constructs that are simple, yet important for characterizing the sample of schools used for the study:

- School size (number of students and grades)
- Average class size (number of students per classroom)
- Location in a rural or urban area
- Size of school catchment area
- Size of school cafeteria relative to student population
- School policies
  - Location of breakfast (classroom or cafeteria)
  - Length and timing of recess and physical education class
  - Time for breakfast and start of school day
  - Busing/student transportation policies

Other important school characteristics are the makeup of the student body in terms of the student and family background variables mentioned above. These features can be estimated using individual level data, or, if only data from sample members are available, they can be requested separately of the school district.

In summary, the evaluation design will include detailed measures of student and family background as well as relevant school and teacher characteristics. These data will be critical for gaining precision in the experimental impact estimates and for modeling the participation decision in the nonexperimental analysis proposed later in this report.

V. RANDOM ASSIGNMENT DESIGN

A variety of impact evaluation alternatives are potentially available for the Universal-free School Breakfast Program (USBP) demonstration evaluation. They vary considerably in validity, complexity, and, most important, cost. Given that the main objective of the Food and Nutrition Service (FNS) is to obtain the most reliable estimates of the impact of the USBP on student outcomes, the preferred approach, and the one developed in this chapter, is an experimental, or randomized, design.

A critical issue that the experimental design must confront is to detect relatively modest impacts in student achievement, given the funding constraints of the demonstration and evaluation. It is almost certain that the nutritional quality of school breakfasts under the USBP and regular School Breakfast Program (SBP) will be essentially the same. If that is true, then the USBP pilot projects will substantially affect only the dietary intake and academic achievement of children who are *new* participants (that is, children who were not participating in the regular School Breakfast Program prior to demonstration implementation but who would begin participating under the USBP).

Under an experimental design that randomizes schools, the impact of the USBP on student achievement will be estimated as the regression-adjusted difference in mean outcomes between the treatment (USBP) and control (SBP) groups of sampled students. These student samples will include both children who participate and those who chose not to participate in the school breakfast program offered in their schools. Since it is anticipated that participation in the school breakfast program at USBP demonstration schools will increase by approximately 25 percentage points, the effect on the children directly affected will be substantially diluted by the children (75 percent) who

experience no effect. This implies that large samples of schools will be needed to detect impacts--possibly larger samples than are feasible, given the $13 million funding constraint.

This chapter develops an evaluation design that represents the best possible approach for detecting USBP impacts, given the funds available for the demonstration and evaluation. The approach is based on a design that pairs schools within districts, then randomly assigns each school in the pair, into USBP and regular SBP status, and finally, selects samples of students from these schools.

The experimental design that is developed must be tailored to the schools participating in the demonstration. However, since the school districts that will apply for the USBP demonstration are unknown at this time, it is not yet possible to finalize all aspects of the design. Thus, in this chapter variants of the basic design approach are presented. The options presented vary in terms of the method used for collecting standardized student achievement test data. If more districts apply than are needed for the evaluation, and most that apply conduct standardized achievement tests of students in which the tests meet evaluation requirements, then the evaluation contractor would not need to administer new achievement tests in all districts. This would free up resources for the evaluation, enabling it to include more schools; this would increase the power of the evaluation to detect USBP impacts on students. Thus, one version assumes that the evaluation will use the achievement data available from district-administered tests; the other assumes that district-administered tests can be used to measure student achievement at baseline, but the evaluation contractor would administer a new achievement test specially designed for the evaluation during the follow-up period.

In addition, because there is uncertainty at this time whether the demonstration will start in September 2000, or later, in January 2001, the design also includes an option to conduct a pre-

implementation survey of students' parents. This survey would be conducted only if demonstration implementation is delayed until January 2001.

The remainder of this chapter presents the evaluation design approach for detecting USBP impacts. Section A summarizes overall design approaches available to evaluate the USBP pilot projects. It concludes that the preferred approach is an experimental design. Section B describes the basic features of the recommended experimental design approach. Section C describes statistical precision and power for detecting USBP impacts on students under the design options.

A. THE CASE FOR AN EXPERIMENTAL DESIGN

The central objective of the USBP evaluation is to determine whether making free breakfasts available to *all* students (1) increases participation in the SBP, (2) increases the number of students who eat a nutritious breakfast, and (3) improves cognitive and school performance and related outcomes of participating students in settings where the demonstration takes place. The challenge for the evaluation is to isolate changes in outcomes due to the USBP program from changes due to other factors.[1]

The ideal method for separating program influences from other factors is to compare outcomes for a set of students who are offered the USBP and the outcomes of the same set of individuals if they were not offered the program. However, once individuals are offered the USBP, it is not possible to know what their outcomes would have been if they were not given the opportunity to participate. It can be approximated only by comparing their outcomes to those of some other group. This other group, denoted the counterfactual, could be the same individuals before the program was

[1]Another objective of the evaluation is to determine how participation in the school breakfast program, in the sense of actually consuming a school breakfast, affects student outcomes. This can be examined within the experimental design, using methods discussed in some detail in Chapter VII.

implemented (the pre-post method), a group similar to the program group in terms of observable characteristics (the comparison group method), or a group that is similar to the program group in terms of both observable and unobservable characteristics (the experimental, or random assignment, method).

In the evaluation literature, methods that yield correct inferences about whether a program improves outcomes in the settings where it is tested are referred to as "internally valid."[2] The rest of this section uses this concept to assess the three design options available: (1) the experimental, or random assignment, design; (2) the pre-post design; and (3) the comparison group design.

1. Random Assignment Designs

Under experimental designs, "units"--in this case, schools--would be randomly assigned to one of two groups: the program group and the nonprogram group. When this approach is implemented correctly, the nonprogram group is statistically equivalent to the program group in terms of observable and unobservable characteristics.

Strengths. It is generally accepted that random assignment methods have the greatest degree of internal validity. The simple and mechanical nature of random assignment is the source of its internal validity. When implemented carefully, the random assignment design ensures that there are no *systematic* differences between the treatment group that is offered a program and a control group that is not offered a program, except the difference resulting from implementation of the program. The effects of program availability are equivalent to postprogram differences between treatment

---

[2]Another important question that an evaluation may wish to address is whether demonstration policies have similar effects in other settings or on circumstances beyond the demonstration school districts. Methods that yield results that can be generalized to other settings are termed "externally valid." However, as noted in Chapter II, Section D, since FNS's objective is internal validity, the discussion focuses on internal validity.

group members and control group members. All other sources of differences are removed by random assignment. Variation will exist because of sampling error, but statistical tests can be used to assess whether the differences in outcomes for the two groups can be attributed to chance. If the tests show that the differences are unlikely to have occurred by chance, they can be attributed to the program.

Weaknesses. Random assignment methods have some weaknesses in the context of evaluating social programs. The designs require eligible participants not to receive program services, so that a control group can be formed. Withholding benefits or services from eligible participants often makes the experimental designs harder to implement. In the USBP demonstration, the benefit is not having to pay to receive a school meal. Control schools would not be able to offer breakfasts free to all students. Compared to most random assignment studies, the services denied to the control schools and students are not substantial. Students in control schools would not be precluded from receiving school breakfasts; it is just that some students (reduced-price and full-price students) would be required to pay some amount toward the meals, as they currently do.

2. Nonexperimental Design Alternatives

Random assignment designs have desirable features. Alternatives to random assignment are pre-post and comparison group designs. However, these designs would provide unclear answers to most of the important questions in the USBP demonstration.

a. Pre-Post Designs

A pre-post design for the USBP evaluation would involve a comparison of SBP participation and student outcomes between periods before and after implementation of a USBP. Under this

design, the group of students being compared with the USBP group is the same group of students before the program was implemented.

Strengths. Given that the assumptions underlying the pre-post design are met (substantial effects occur within a short period of time), then one important advantage of the designs is that their implementation is relatively simple and straightforward.[3] The designs also are easy to understand because the resulting data have an intuitive meaning: if outcomes after program implementation are better than outcomes before program implementation, it is a straightforward conclusion that the better outcomes are due to the program.

Weaknesses. Few social programs have dramatic effects in a short time. The norm is longer-term followup to assess whether programs have had moderate effects. The basic problem with a pre-post design is that it risks making misleading inferences about program impacts, since any changes that are observed may be due to factors other than implementation of the USBP, especially if substantial time has elapsed between the pre-program and post-program data collection points, or if participants come to programs after they experience adverse events. For example, suppose one observes that participation is higher in the post-implementation than the pre-implementation period for low-income students eligible under the regular SBP. One would like to conclude that this increase in participation among low-income students is due to implementation of the USBP. However, the increase might be due to other factors, such as a downturn in the economy, which could increase program participation.

Another change that could affect a pre-post analysis of the USBP is changes in student composition from year to year. The importance of this factor would be minimized if the USBP were

---

[3]Except, possibly, for timing issues associated with collecting pre-implementation data, as discussed below.

implemented in the middle of a school year, so that pre- and post-implementation variables were measured during the same school year. A more realistic implementation process would imply implementing the USBP at the start of a school year. Also, it is possible that the breakfast-eating patterns of students are largely formed early in the year and that the full effects of a USBP would not be observed in a six-month observation period beginning in midyear.

For these reasons, pre-post designs are probably not well suited to the USBP demonstration.

b. Comparison Group Designs

Conceptually, the nearest alternative to random assignment is the comparison-group design. In comparison group designs, a comparison group typically is formed by identifying individuals who are similar to individuals exposed to a program but who, for whatever reason, do not participate. For the USBP demonstration, comparison group designs would involve comparisons of data collected at the same time for students who are and are not in schools offering the USBP program. A comparison group design would use comparison schools (either within or external to the USBP school districts) and compare outcomes of students in schools that choose to implement the USBP with outcomes of students in schools that do not implement the program. Another strategy is to identify a comparison group from a national database.

Strengths. Comparison group designs improve on pre-post designs in that the comparison group creates useful benchmarks for assessing change. The comparison group's outcomes reflect the influence of time and maturity, allowing the evaluation to provide a clearer look at whether the treatment group's outcomes are being influenced by the program. In some circumstances, comparison group designs may also be easier to implement than random assignment designs.

It should be noted that these strengths do not mean that comparison group designs have lower implementation and data collection costs than random assignment evaluations. In fact, comparison

group designs generally are more expensive than random assignment designs precisely because a comparison group must be recruited, whereas the control group is readily identifiable in experimental designs.

Weaknesses. The major weakness of comparison designs is that observed differences in outcomes are not necessarily program effects--the design may lack internal validity. Suppose the USBP were implemented in a set of schools, and outcomes of participants in the program were compared to outcomes of students in other schools to assess program impacts (where comparison schools are either within or external to the USBP districts). Differences among schools that choose the USBP program and those that do not could themselves affect the comparisons. For example, School Food Authorities (SFAs) selected for the demonstration might choose to implement the program in the elementary schools with low participation rates among students eligible for the regular SBP, in hopes that a universal program would substantially improve participation rates. In this scenario, a key outcome variable would differ systematically between the schools that could be used for comparison purposes and the schools implementing the USBP. Whereas many such differences can be measured, and potentially controlled for, in the analysis, some differences will remain unmeasured. If these unmeasured differences affect program outcomes, there is no direct, reliable way to separate their effects from program outcomes.

It is not possible to know in advance whether unobserved factors cause comparison group outcomes to be higher or lower than treatment group outcomes. It is safe to say, however, that evaluation results will be affected by at least some bias when comparison group designs are used. There are strategies for reducing the bias. For example, an evaluation using this design could create different comparison groups that may incorporate different types of bias. It would then average the results to reduce the biases. Also, the use of pre-post measures of student outcomes can partly

normalize for differences between groups of students and reduce bias. These approaches may improve the validity of a comparison group design, but they would also increase data collection costs and evaluation complexity.

3. The Importance of Obtaining Reliable Evaluation Results Favors Using a Random Assignment Design to Evaluate the USBP Pilot Projects

Trade-offs exist in any program evaluation, more so when schools are the setting for the program. When the pros and cons of available evaluation design alternatives are considered, the decision boils down to whether to use a comparison group design or a random assignment design in evaluating the USBP pilot projects. Random assignment designs are clearly superior, yielding internally valid impact estimates. Comparison group designs are plagued by the high likelihood of yielding invalid results.

The fundamental question to ask is: Is the invalidity from using a comparison group design within a tolerable range in the case of the USBP demonstration? The answer to this question depends on the value of the information. Expanding the current school breakfast program to one in which breakfasts are free to all students, regardless of family income, may substantially increase the cost of the school breakfast program. In an economic climate where public resources are constrained, it is critical to obtain strong information on whether this type of expenditure is worthwhile--whether it increases the number of students consuming nutritious breakfasts, improves dietary intake, and enhances student cognition and academic performance.

It is hard to argue that little is riding on the outcomes of USBP programs. Random assignment designs are more likely to withstand the intense scrutiny an evaluation of the USBP will receive. Given the costs of implementing the USBP nationally, being "close enough" probably is not considered acceptable.

Given FNS's objective of obtaining reliable findings, random assignment's clear superiority in this area over other design alternatives favors using random assignment designs for the USBP evaluation.

The rest of this chapter specifies the features of the random assignment design.

B. FEATURES OF THE RANDOM ASSIGNMENT DESIGN

A number of decisions need to be made when developing a random assignment evaluation design for the USBP demonstration. This section describes overall features of the design approach for the random assignment impact study. These include decisions about the level at which randomization would occur, whether to formally incorporate planned variations in the treatment, and the definition of the counterfactual (control) group. Decisions about the number of districts and schools to work with and the precision of the estimates, are covered in Section V.C.

1. Preferred Approach Is to Randomize Schools

In principle, random assignment can be imposed at a variety of levels. It could be implemented at the school level; or, within schools, students or classrooms could be randomly assigned. An issue to be resolved is, at what level in the USBP demonstration will random assignment be administered.

In the design of random assignment evaluation, it is often useful first to examine the implications of random assignment at the lowest level--in this case, with students--and then to consider whether the program's structure will be violated or the evaluation's results flawed by such a design. If violated or flawed, then one needs to consider doing random assignment at the next higher level (in this case, classrooms) and repeat the exercise, stopping when the choice does not violate the program's structure and yields unbiased results.

Using this process, the preferred approach is to randomly assign schools, not students or classes. If students are randomly assigned within schools, the effects of the USBP on participation rates (and dietary intake) may not be representative of a fully implemented program, thus lessening the value of the demonstration. If random assignment were used to offer the USBP to a random subset of students within demonstration schools, these students would constitute the treatment group, and the students not offered the program would constitute the control group. Comparison of outcomes of the two groups potentially could yield unbiased estimates of program impacts. However, a possible serious problem with randomizing students instead of schools is that a USBP implemented in only part of a school probably would not have the same effects on participation as a program implemented in the entire school. As a result, it may not fully address some of the evaluation's most critical research questions. The beneficial effects of the USBP on participation and other outcomes are achieved, in part, because all students can receive a breakfast free, thus reducing the effects of stigma. Randomizing students within a school defeats this purpose, since only some students would have an opportunity to have free breakfasts, whereas others would not. In addition, a small demonstration program might not be able to offer the same kind of breakfast or be publicized in the same way as a fully implemented program. Within-school randomization of students also is likely to be more complex to implement than randomizing schools, so that school districts and schools may resist the implementation. Similar arguments hold for why it is not appropriate to randomize classes within schools.

Using random assignment to choose which schools in participating SFAs offer the USBP eliminates biases from self-selection of students and would be relatively easy to implement. However, random differences between schools and their experiences over time, all else constant, will make it more difficult to detect program impacts. The loss of precision arises because sample

members in a group (students in treatment and control schools) are affected by common events or circumstances. This grouping causes the estimator of the program impact, the average outcome difference, to vary more than when students are randomly assigned.

In practical terms, the implication of having schools as the units of intervention is that, to estimate program effects accurately, the evaluation will need to include more schools. Increasing the number of schools in the treatment group and the control group reduces the variability of the impact estimator, because random events affecting schools have a better chance of canceling each other out. In addition, variation across schools can, to some degree, be mitigated by matching pairs of similar schools within each SFA, then using random assignment to choose which school within each matched pair could offer the USBP.

Matching and Randomizing Schools. Within each SFA, similar schools will be matched into pairs, then random assignment will determine which school within each matched pair can implement the USBP. This process should result in a random sample of treatment and control schools that are matched on a set of observable criteria, thereby reducing the cross-school variation in outcomes.

To accomplish the matching and randomizing of schools, the first step is to consider what criteria to use in the initial pairing. Several important characteristics to consider are the following: number of students enrolled, percentage of students certified for free- or reduced-price meals, racial and ethnic composition of the school, the SBP participation rate, and school-wide average achievement test scores. The first three criteria--school enrollment, the percentage of students certified for free- or reduced-price meals, and race and ethnicity--are important for ensuring that (1) treatment and control schools are similar on demographic and socioeconomic characteristics, and (2) the face validity of the matching process. Given that both program participation and school achievement are key outcomes for the evaluation, using these variables as matching criteria helps

make the treatment and control sample comparable on key study outcomes prior to USBP implementation.

Based on experience on other studies using matched comparison areas, sometimes it is important to limit the number of criteria used in the matching process. The main reason for this is that using too many criteria for matching schools prior to randomization may lead to pairs of schools that are not similar on the most important characteristics.[4] For example, if the most important matching criterion is, say, student achievement scores, then using the other criteria described above may lead to pairings of schools that are less well-matched on achievement scores.

The exact matching criteria are, in some sense, judgmental. Given the importance of academic achievement, it is likely that this criterion should be used to pair schools before randomization. The other criteria described above also are important for ensuring face validity. In addition, working with the SFAs and school districts is important for ensuring that similar schools are paired before randomization. Often, school district staff are in the best position to offer suggestions on the schools in their district that are the most similar. The process would be iterative: the evaluation team will make initial pairings based on administrative data, then revise them based on input received after the pairings are reviewed by school district officials.

---

[4]The evaluation of the infant mortality initiative, Healthy Start, for example, was based on a comparison site design in which outcomes for the demonstration projects were compared with outcomes for matched comparison areas. The selection of comparison areas first used an extensive set of matching variables, but the selected comparison sites were not well-matched to the demonstration projects on the key outcome of infant mortality. After considerable exploratory analysis and consultation with the project staff, the matching criteria used were only baseline infant mortality rates and race and ethnic composition.

2. The Design Does Not Include Planned Variation in the Demonstration Treatment

Given the decentralized nature of the regular SBP, SFAs and schools in the demonstration will likely implement the USBP in a number of different ways. Some schools will simply make all breakfasts free and operate the program just as the SBP did. Other schools may change where the breakfasts are provided; for example, they might offer the breakfasts in classrooms rather than the cafeteria, or they might send all children to the cafeteria as they arrive at school. SFAs may also be creative about the menu items offered, or they serve hot meals.

Ideally, the evaluation would determine which program features are effective and which types of students are affected. However, learning what works within programs is both difficult and expensive. Schools would need to be grouped, and then randomly assigned to different treatment settings or a control group, since randomizing schools in this way would be the only statistically reliable way to assess the importance of various program alternatives and why they are effective.

The possibility of designing a demonstration that explicitly creates variation in treatments is limited in the current situation. To achieve reasonable levels of precision would require considerably more resources than are available for the demonstration evaluation (see the sample sizes shown in Section V.C, for the treatment group multiplied by the number of different treatment interventions). In addition, because of increased complexities, implementing this design would require greater resources. To implement the design, the evaluation contractor would need to carefully define the program variations up front, as well as randomly assigning schools to one or another of these variations. That would require considerably more discussion with school districts than would be required if there were a single treatment--to get them to understand the process and then stick to their assignment. This would require substantial resources, and seems to go beyond the congressional mandate.

Thus, given resource constraints, it is not feasible to plan variation in the treatment in order to assess the impacts of different program types. Furthermore, when offering the USBP, it does not seem feasible to constrain SFAs and schools to adoption of a single type of program setting. Given the current decentralization of decision making in school districts under the regular SBP, and based on information from Child Nutrition Program administrators, under a national USBP policy, SFAs and schools probably would be free to provide breakfasts in a range of settings, such as school cafeterias and classrooms. Since that is how the program would be implemented in the long run, from that standpoint it would be more desirable to test a combination of treatment settings than a single one. That is, demonstration SFAs and schools should be allowed to implement whatever program they view as sensible, given a school's normal breakfast program and student needs and preferences--subject to meeting the regulations and demonstration requirements of the USDA breakfast program.

Information about the effectiveness of variations in program features is important for program design, however. Since the demonstration will not explicitly incorporate treatment variation into the design, it is important when developing the analysis plans to consider evaluation approaches that will at least shed light on program effectiveness. Here, the role of the process analysis of program implementation and operation is crucial (see the discussion in Chapter VIII). Other strategies will, where statistical precision constraints permit, include disaggregating parts of the impact analysis by type of demonstration treatment, explicitly controlling for treatment variation when analyzing data pooled across different interventions; and conducting descriptive analyses of differences in student outcomes for different treatment variations (the latter will be useful even when available statistical precision is relative low, so as to at least determine whether there is any hint of evidence of major treatment effects on outcomes).

3. The Regular SBP as the Counterfactual

Identifying the appropriate group to compare with the treatment group is another of several key elements of the evaluation design that needs to be specified. Deciding on the appropriate counterfactual is important because it defines how the evaluation results can be interpreted. There are three potential counterfactuals for the USBP evaluation: the regular SBP, the no-breakfast program, and a counterfactual that includes both schools with the SBP and without it.

Having the regular SBP as the counterfactual means that the evaluation results will be interpreted as indicating what would happen if the SBP were replaced with a USBP. That is, it indicates how participation in the SBP would change and how dietary and school-based outcomes would change, *relative* to an environment in which the SBP is offered to low-income students. For example, changes in dietary intake might occur because the dietary intake of students who participate in the USBP, but who would not participate in the SBP, may change. Key outcomes might also change for students who would participate under either program, if the USBP changed the breakfasts that were offered due, say, to an increased scale of program operations.

If the counterfactual were the no-breakfast-program option, the evaluation results would be interpreted as answering the more basic question of whether the provision of a breakfast program affects student dietary and school-based outcomes. This comparison would also make possible an assessment of the effects of the USBP on participation.

FNS has provided direction on the decision about which counterfactual is appropriate for the study. It recommends that the demonstration be limited to those schools already participating in the SBP, making SBP schools the counterfactual for the demonstration evaluation. While a design that includes schools with no breakfast programs, a regular SBP, and a universal-free breakfast program has some appeal, several considerations point toward using only the SBP as the counterfactual:

- ***The majority of schools currently offer the SBP.*** More than 75 percent of public schools currently offer the SBP. This high participation rate suggests that schools that do not participate in the current program are likely to differ substantially from those that do participate. The evaluation would need to control for those factors, which adds complexity to the design. In addition, although one could find schools without an SBP in which to implement a USBP, the likelihood of having an adequate number of schools not on the program is relatively low.

- ***The legislation authorizing the demonstration appears to pose the SBP as the counterfactual.*** Section 109(b)(i) states that the goals of the pilot projects are (1) to "reduce paperwork, simplify meal counting requirements, and make changes that will *increase participation in the school breakfast program*, and (2) to evaluate the *effect of providing free breakfasts to elementary school children, without regard to family income*, on participation, academic achievement, attendance and tardiness, and dietary intake over the course of a day" (italics added). Congress recognizes that there are specific reasons why some schools currently are not participating in the SBP, such as transportation or scheduling issues or the fact that the district is too affluent to make the program cost-effective. The issue is whether a policy of moving schools currently participating in the SBP toward universal-free breakfast programs is an appropriate policy change.

- ***It is too costly to include both no-breakfast and regular SBP as counterfactuals.*** Although the evaluation design could, in principle, incorporate both counterfactuals, this approach might be too costly. If the USBP were implemented both in schools with and without the SBP, it would be possible to combine the two alternatives into a "does not offer USBP" counterfactual. However, it is likely that the two types of comparisons would need to be done separately. This would dilute the ability to detect program impacts, particularly when the number of schools in the study is likely to be constrained for budget reasons and because the legislation restricts the evaluation to six SFAs.

4. The Evaluation Will Collect School-Level and Student-Level Data

The evaluation questions posed in the authorizing legislation focus on the effect of offering a USBP on student decisions to participate in a school breakfast program and the effect of that participation on various student outcomes.[5] This focus suggests that the principal unit of analysis will be individual students. However, it is important to note that, even though the ultimate effects of interest occur largely at the *student* level, in some instances it will be easier and cheaper to obtain

---

[5]The legislation authorizing the demonstration lists a school-based question as well, which is to measure the effect of the USBP on the paperwork schools are required to complete.

average *school-level* data (i.e., school averages of student characteristics and outcomes) on some variables and to collect such data over several points in time.

The evaluation will include both the collection of school records data on school averages and student-level data for sampled students; it will also conduct surveys of sampled students. Some analysis of administrative records data would take place at the school level, and some will take place at the student level. (Chapter VII discusses student- and school-level analyses.) Both units of analysis are important and are included in the evaluation design. For some outcomes--such as SBP participation, student absenteeism, tardiness, disciplinary actions, health (visits to the nurse while at school), and student academic achievement--the evaluation will conduct analyses at both the school level and the individual-student level. For outcomes such as dietary intake and quality of diet, cognition, and student behavior, the evaluation will measure the outcomes by surveying individuals and/or their teachers and parents, since schools typically do not collect the data on individuals, and, because of the costs or logistical problems. In this situation, the appropriate unit of analysis is the individual student.

5. The Evaluation Will Exclude Kindergarten Students from the Survey Research Sample

The legislation authorizing the demonstration restricts participating schools to elementary schools. The growing trend in the United States is for elementary schools to include kindergarten and grades 1 through 5 (although some elementary schools include grade 6). Even if all the aforementioned grades are included in the demonstration, there is the issue of whether the evaluation should include all the grades.

Of particular concern is whether the evaluation should include kindergarten students. Including kindergartners introduces several complications. Kindergarten is not universal, the way other grades are. Many kindergartens operate on half-day schedules and thus do not provide a school breakfast

program for kindergarten students. Moreover, children attending afternoon kindergarten sessions at schools that have breakfast programs would not get the opportunity to participate in a school breakfast program. One solution would be to confine the analysis to kindergartens that had morning operations (that is, those that include full-day kindergartens and those with morning half-day kindergartens, but excluding afternoon-only programs). These analyses, however, would need to carefully control for type of kindergarten (full-day versus morning session only) and for other student and family characteristics, since children attending full-day and morning-only kindergartens probably differ. If the alternative--including all kindergartens--were adopted, then the evaluation would need to impose a different sampling strategy for afternoon kindergarten children. This is because family and student characteristics may differ among morning and afternoon children, and it could not be assumed that data collected for morning students would be representative of afternoon or all-day kindergarten students.[6]

In addition, some of the key outcome measures would be difficult to collect from kindergarten children. For example, many dietary surveys have not been administered to these children. The elementary school component of the SNDA-1 sample included children in grades 1 to 5. The 1994-1996 CSFII dietary recall instrument was administered directly to children 6 and older (with some assistance from parents of children 6 to 11), but parents completed the instrument for children younger than 6 (which would include most kindergarten children). Published achievement tests are available for grade K, but districts and states are generally reluctant to use them, preferring instead to use "readiness tests." Usually, special-education status is not adequately assessed before first grade--a key control variable that would not be available for kindergarten students.

---

[6]There are alternatives to public school kindergarten programs, such as Head Start and private programs. This complicates the analyses, since it means that the USBP may affect whether or not students attend public school programs, possibly affecting who shows up in public schools.

The above discussion indicates that several complications arise from including children in kindergarten. While these complications are not insurmountable, the added complexity, as well as the uncertainty about the comparability of findings between kindergartners in treatment and control schools, leads to a recommendation to exclude kindergarten children from the survey research samples, but not from other components of the evaluation. For example, using available administrative records data, the evaluation will examine the effects of the USBP on these children's participation and school attendance and other outcomes.

6. The Evaluation Will Include a Longitudinal Component

An important design question to resolve is: How many rounds of outcomes data should the evaluation collect on dietary intake and student achievement? There are two issues to take into consideration when making this decision. The first is whether multiple rounds of student outcomes data are required in order to reliably estimate USBP impacts on students and, if so, how many. That is the focus of this section. The other issue is how *long* students should be followed to assess both short- and longer-run program impacts, which is discussed in the next section.

It is usually not necessary to supplement follow-up data on treatment and control group outcomes with the collection of baseline *outcomes* data to reliably estimate program impacts. If random assignment is performed correctly, then treatment and control members will not differ systematically in terms of measured or unmeasured characteristics at baseline, regardless of whether the characteristics are outcomes or not; therefore, simple comparisons of the mean outcomes of treatment and control groups during a *post-implementation* (follow-up) period will produce unbiased estimates of program impacts.

More often than not, baseline data on characteristics (including data on some outcomes) are collected, since collecting these data usually is straightforward and relatively inexpensive and

benefits the analyses in the following ways: it is possible to improve the precision of the impact estimates by using multivariate analytic models to control for the characteristics of individuals (particularly useful when conducting subgroup analyses); and it results in both more characteristics in which to evaluate whether random assignment was implemented correctly and to consider how impacts vary by subgroup.

Deciding whether the USBP evaluation should collect student outcomes data at baseline for outcomes such as dietary intake and student achievement (if the evaluation contractor has to administer achievement tests to students) is especially important in the USBP demonstration evaluation because of the relatively high costs involved in collecting the data. Administering dietary intake interviews and conducting achievement tests is extremely expensive because of the complexities of the data collection and the need to collect the data in person. Given evaluation resource constraints, collecting baseline data on dietary intake and achievement mean that follow-up data collection on other types of outcomes must be limited to a single round of data collection. To justify that kind of trade-off, there must be vital analytical reasons for collecting outcomes data at baseline.

It turns out that it is critical to obtain information on student achievement at two points in time (either at baseline and at followup or at two follow-up periods).[7] Even with relatively large samples and a randomly assigned control group, it is expected that the evaluation will not be able to estimate student achievement reliably if based only on a single followup, because of the need to detect relatively small impacts in the overall student sample (see Section V.C). The variation in test scores due to student abilities and the variation due to the myriad differences in education policies and

[7]Having achievement test data for three points in time would be even better, since it would enable the evaluation to implement an individual, fixed-effects estimation method that would control to some degree for selectivity bias (see discussion in Chapter VII).

interventions across schools and districts will make it difficult to detect any effects of USBP availability. The available evidence on the determinants of achievement test scores suggests that, at a minimum, one would need longitudinal data with two time points (a benchmark and followup) for each sampled student; the evaluation should also collect data on family background and student disability status to control for those characteristics.

The evaluation will obtain baseline and follow-up data on student achievement. (Whether the data can come from existing district-administered tests or new tests to be conducted by the evaluation contractor cannot be decided until FNS finds out which SFAs apply.) In reliably estimating impacts, it appears less critical to collect two rounds of data on other student outcomes, such as dietary intake. Analyses of CSFII data suggest that dietary intake at any point in time is weakly correlated to prior intake, and the correlation becomes substantially weaker as more time elapses between preprogram and postprogram data collection points.[8] The analytic benefits of having data on dietary intake at two points in time (reducing variance) do not appear to justify the increase in costs of collecting the data.

The Cross-Sectional and Longitudinal Samples. The evaluation will include two partially overlapping samples. One component is a cross-section sample of students from all grades 1 through 6.[9] If the demonstration starts in school year 2000 - 2001, then the cross sectional sample will be students in grades 1 through 6 in SY 2000 - 2001 (see Figure V.1). These students would be

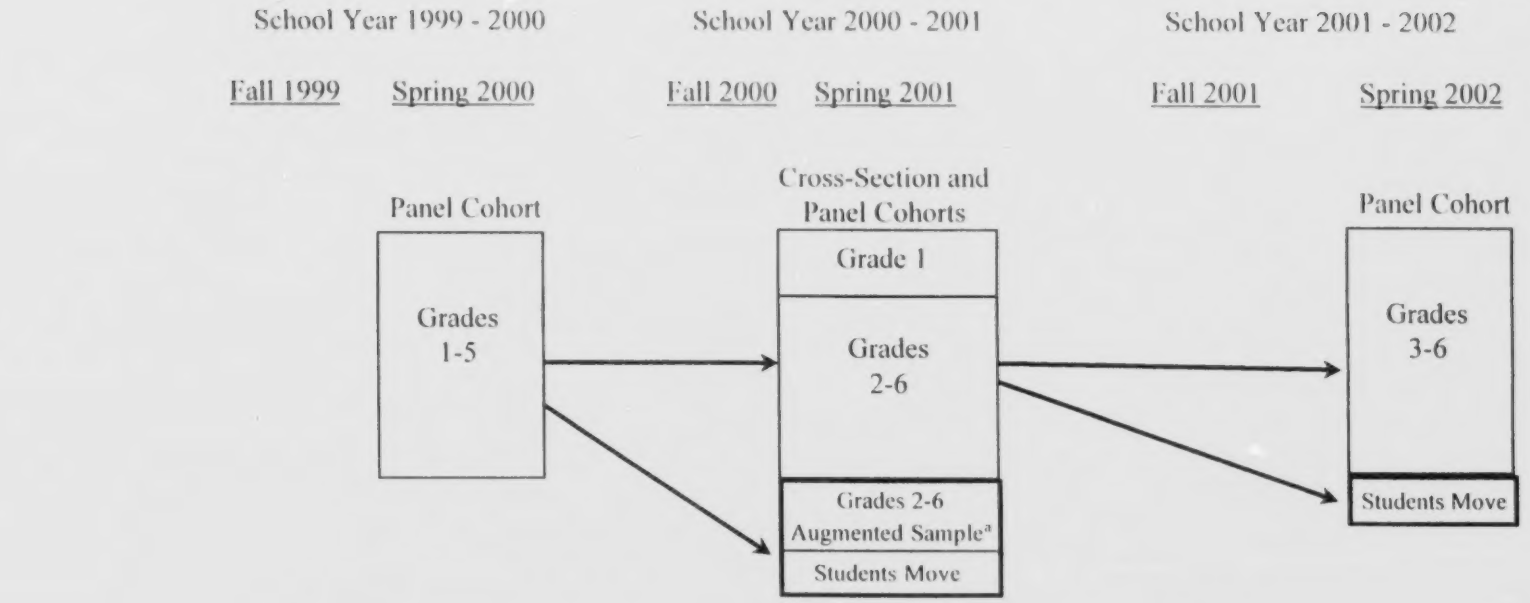
---

[8]The CSFII collects dietary intake data on individuals at two points in time. We regressed CSFII sample members' intake in Day 2 on a variety of control variables, first without Day 1 intake, then with Day 1 intake. We examined whether the earlier intake data were highly correlated with later intake, resulting in a huge increase in the R-square. The R-square increased, but not substantially, and the overall R-square remained fairly low.

[9]Students in Kindergarten will also be included, but only in the analyses of administrative records data (see the earlier discussion in Section V.B.5).

FIGURE V.1

DATA COLLECTION DESIGN: STUDENT SAMPLES



= Panel Cohort

surveyed once, during spring 2001. At that time, sampled students would be administered dietary recalls and given cognition tests; they would also be surveyed about their attitudes toward school breakfast and school climate. One of their parents would also be administered a short survey to obtain data on parental attitudes and household characteristics, including the household's food security.

The second component tracks a longitudinal sample of cohorts of students in the study schools, based on administrative records.[10] In its most basic form, the longitudinal sample would consist of the cohorts of students in grades 2 through 6 at the time of demonstration startup in SY 2000 - 2001 (see Figure V.1). The baseline measure of student achievement would come from district-administered tests taken a year earlier, in spring 2000, when the students were in grades 1 through 5; the follow-up measure would be from either a district- or evaluation contractor-administered test, taken in spring 2001, when the cohort of students are in grades 2 through 6.[11] Operationally, the sampling for the longitudinal analysis entails selecting a somewhat larger cross-sectional sample of students in grades 2 through 6 in SY 2000 - 2001, then otherwise would be needed to support the planned cross-sectional analyses. The larger sample is necessary to offset the influence of student mobility. That is, since students moving into the school district during SY 2000 - 2001 would not

---

[10]As noted, the longitudinal data collection would involve one round of achievement testing administered by the evaluation contractor, if district-administered test data are not available for the follow-up period.

[11]If the districts test elementary students annually so that three years of achievement test data are available, then the evaluation could estimate impacts on student achievement using an individual fixed-effects "difference-of-differences model" (see discussion in Chapter VII). In this case, the cohort of students would be students in grades 2 through 5 at the time of demonstration startup. The three student achievement measures would be: one in the year before implementation of the USBP (spring 2000), when the student cohorts are in grades 1 through 4; one during the first year of the demonstration (spring 2001), when students are in grades 2 through 5; and one in the second year of the demonstration (spring 2002), when students are in grades 3 through 6.

have an achievement test score for the prior school year, SY 1999 - 2000, proportionately more students in grades 2 through 6 need to be sampled for the evaluation. It is anticipated that approximately 15 to 20 percent of students in the demonstration sites will move into the school the district since the spring of the previous year.

Following individual cohorts of students for at least two points at a time is the best design for assessing program impacts on student achievement. Which grades to follow will ultimately depend on the schools selected for the demonstration and which grades and the frequency students are tested. Interviewing students from the full set of grades 1 through 6, who, at the time of data collection, attend study schools allows for an overall analysis of the effects of the school-level intervention on students attending the schools.

Attrition. Most likely, all or nearly all the schools selected for the evaluation will remain in the study sample for the duration of the evaluation. Students, however, move, either within school districts or out of school districts. Student mobility often is a major analytic issue to consider in large-scale evaluations, especially with interventions known as "intention to treat." Such interventions often use the individual student as the unit of assignment and analysis and, when mobility occurs, it is critical to follow those students who move to assess impacts on the original group "offered" the intervention. Tracking students who move is costly and difficult. Ultimately, study findings may be affected by the extent of mobility and the ability to track students who move.

Many of the planned data collection activities--dietary recall, student and parent surveys, cognitive tests, and teacher surveys--are single point-of-time efforts, so student mobility is not a major issue. The collection of student achievement data, however, introduces a longitudinal component to the USBP evaluation. Because of the need to examine the impact of school breakfasts on growth in student achievement, achievement needs to be measured at a minimum of two points

in time. Sample attrition is an issue to consider when planning the collection of student achievement data.

Fortunately, because schools are the recommended unit of assignment for the pilot projects, the mobility of students in the context of the USBP demonstration and evaluation does not have the analytic complications (other than the need for higher initial sample sizes) it would have if individual students were the unit of assignment. Unless student mobility is related to the USBP intervention (which is unlikely), there is no analytic reason to follow those students who move out of the study schools and who do not receive the USBP intervention during the period of time when growth in achievement is measured.[12]

The main implication of student mobility is that the sample used to examine program impacts on student achievement must be students who are in the study schools at the two points in time when the achievement data are collected. Because of the expected attrition between school years, initial sample sizes for the student achievement study component must be large enough to account for the fact that a certain percentage of those sampled will move and not have follow-up data on student achievement. Alternatively, the student achievement sample could include some students who move to the USBP schools during the second year of the demonstration, if achievement test data could be obtained from their school records for the previous year.

7. Length of the Follow-up Period

Another aspect of the design that must be specified is the length of the follow-up period. Ideally, one would like a follow-up period long enough to assess both short- and longer-run impacts

---

[12]For students who were initially in the USBP schools, if they move from those schools and are not there during the time when change in student achievement is being measured, there can be no impact of the USBP for them. If the USBP were implemented nationwide, presumably students who move would move to other schools with the USBP.

of the USBP on elementary school students' academic achievement. The legislative requirement that the demonstration is to last for a period of three consecutive school years means that students at treatment schools will, at most, be able to receive breakfasts in a universally free setting for three years; after that, they will revert to the regular SBP program, should the schools they attend participate in it. The critical question to answer in the case of the USBP demonstration (since it lasts only three years) is whether the evaluation should conduct longitudinal followup on students during periods after the demonstration ends, in order to assess longer-run impacts.

Alternative 1: Confine Student Followup to the Demonstration Period. One possible design would limit student followup to the period in which the demonstration is active--that is, when students are offered USBP meals. Such a design could support a three-year longitudinal followup for students initially in grades 1 through 4 (if elementary schools in the study include grade 6). Impacts on students could be assessed at the end of year 1, then in the second year, and, resources permitting, during the third year of the follow-up period.

The advantage of this design is that it provides information on the effects of the USBP on student outcomes while it is still possible to participate in the program. For example, the design could find out whether the program has a different impact on students who start receiving USBP meals in the first grade and stay through the third grade, than it has for students who begin participating later, say, in the third grade and continue receiving meals through the fifth grade. In addition to assessing the impacts according to the timing of participation, the evaluation could examine the impact of participation according to the duration of participation (for example, received meals for one year versus two years versus three years). In principle, the effects of the duration and

timing of participation can be estimated from a regression model, with duration of time in the school and grade of entry into the study school as explanatory variables.[13]

Alternative 2: Conduct Student Followup After the Demonstration Ends. An alternative design would include followup during the postdemonstration period, when USBP breakfasts are not available. Such a design would enable the evaluation to address research questions such as: (1) What are the longer-term effects of USBP participation on participation in the regular SBP? What are the longer-term impacts on student achievement? Do effects on these outcomes persist over the longer term, or do they fade over time? Do impacts appear later in the students' academic career? Do they become stronger over time?

Take the case of participation in the regular SBP. Suppose that participation in the regular SBP by low-income students in the postdemonstration period is greater for students initially offered USBP breakfasts than for children attending regular SBP schools. That may occur because, by providing breakfasts without regard to family income, the USBP reduces the stigma of program participation, and that impression is carried forward by students into their middle-school years. Based on the finding, Congress, to reduce the costs of the USBP, might decide to authorize it in only elementary schools, satisfied that participation in the regular SBP will remain high for children most in need as they move into middle and high schools.

The situation is more complex in the case of achievement. Suppose that over a longer period, the impacts on student achievement disappear. What does that mean? It is consistent with no longer having access to the USBP: a positive impact might have persisted if the program were still in place. But it also might be the case that the effect would have disappeared even if the program were still

---

[13]For students attending USBP schools, the control variables would be program-school duration and the grade at which they enter the demonstration. The variables would be defined similarly for students attending control schools (regular SBP), but in terms of control schools.

available to students, reflecting the normal maturing of students. However, without giving students the opportunity to participate in the USBP over the longer period, it is impossible to disentangle the reason.

The main disadvantages of a design that follows students beyond the demonstration period are that it is complex and expensive to implement. While the research objectives that require collecting data after the demonstration appear interesting, they appear to be of second-order importance and the data could not always be interpreted. In addition, unless more resources could be added to the demonstration and evaluation, it is not feasible to collect data beyond the demonstration period. Focusing on the demonstration period appears to be the most prudent use of resources, and will be followed for the USBP evaluation.

C. STATISTICAL PRECISION AND POWER

This section discusses the statistical precision levels for the four design alternatives. Section 1 defines the design alternatives that are considered. Section 2 describes the key assumptions that underlie the precision analysis, while Section 3 highlights a number of factors that substantially influence the likely statistical power levels that can be achieved. Detailed findings about trade-offs between statistical precision and sample sizes for various aspects of the analysis are presented in Section 4.

1. Design Options

The design to evaluate the USBP pilot projects will match and randomize schools. The actual number of schools to be included in the demonstration and evaluation will depend on demonstration school districts' ability to provide achievement test data on sampled students.

Design Alternative 1, the preferred design, assumes that districts will be able to provide data on student achievement test scores at both baseline and during the post-implementation follow-up period. If demonstration school districts cannot provide follow-up data on student achievement, then the evaluation contractor would need to conduct one round of achievement testing during the follow-up period (Design Alternative 2). Because it is more costly to test students then obtain existing test data from demonstration school districts, fewer schools and students would be sampled under Design Alternative 2. The design alternatives are as follows:

- ***Design Alternative 1a: Districts Will Supply Baseline and Follow-up Student Achievement Data.*** The first design alternative pairs 144 schools, then randomly assigns each school in the pair into either treatment (USBP) and control (regular SBP) schools--72 schools in each of the treatment and control groups. It collects detailed data on 30 students per school, or a total of 4,320 students. Under this design, administrative records data and student survey and test data would be used to assess the impact of the USBP on participation and a variety of student outcomes such as dietary intake, attendance, cognition, and student achievement.

- ***Design Alternative 2a: Evaluation Contractor Conducts Student Achievement Test During the Follow-up Period.*** This design pairs 120 schools, then randomly assigns each school in the pair into treatment (USBP) and control (regular SBP) schools--60 schools in each of the treatment and control groups. It would collect data on 30 students per school, or a total of 3,600 students. As with the first design, this design will use administrative records data and student survey and test data to assess the impact of the USBP on participation, dietary intake, attendance, cognition, and student academic achievement.

As discussed later in this chapter, one potential limitation with both designs is that program effects are difficult to detect because of the dilution of effects on "treatment" students who participate by the substantial number of students who do not participate and hence are not directly affected by the program. If implementation of the USBP demonstration should be delayed until January 2001, then a preimplementation survey of sampled students' families in order to identify students who would likely become new USBP participants, could be conducted under either design as a means for

addressing this problem. Each design alternative therefore includes an option for conducting a preimplementation survey (these are denoted Design 1b and 2b, respectively).

For each of the above alternatives, as discussed in Chapter VII, two major types of analysis of outcome variables would be conducted. One method--the analysis of *availability*--essentially compares the means of the students assigned to treatment status and those assigned to control status.[14] It is this first method that fully draws on the experimental design; the experimental design essentially eliminates risks of selection bias problems. However, a second line of analysis will focus direction on *participation*, comparing USBP students who actually participate in the program with those of comparison students not receiving school breakfasts.[15] This analysis is more subject to problems cause by selection bias, but it has greater statistical power if selection bias proves not to be a significant issue.

In the discussion that follows, power levels associated with each of these two lines of analysis are considered.

2. Assumptions

The approach to examining the precision of estimates for proposed sample sizes for the two design alternatives and options is to estimate the smallest *true* program impact that will likely lead the evaluation to conclude statistically that the USBP program has an impact. The calculations assume the following:

---

[14]To increase precision, these and other comparisons discussed in the text will be regression-adjusted. See Chapter VII.

[15]Separate comparisons of USBP participants and USBP nonparticipants, and of USBP participants and SBP participants, will also be conducted.

- ***Using One-Tailed Difference-of-Means Tests at .05 Significance Level, 80 Percent Power.*** The evaluation will conclude that a program has an impact if the statistical test rejects, at the 95 percent confidence level, the hypothesis that treatment and control means are identical. The minimum detectable impact is calculated so that the evaluation has, at a minimum, an 80 percent chance of concluding that the program had an effect if, in fact, an effect of the assumed size exists.[16]

- ***Not Generalizing Beyond the Demonstration SFAs.*** The evaluation findings will generalize to the six SFAs in the demonstration. That is, since the objective of the evaluation is to reliably estimate the impacts of the USBP on students in schools in the demonstration districts (internal validity), the findings do not generalize to all SFAs in the country (that is, the between-SFA variance component is taken out, or ignored, in the calculations).

- ***Balanced Design.*** It is assumed that there is an even split between treatment and control schools in each SFA.[17]

- ***A 25 Percent Increase in School Breakfast Participation.*** It is assumed that the USBP will increase student participation in school breakfast by 25 percentage points (from 30 percent to 55 percent).

- ***Using Reading Score as the Achievement Measure.*** For the analysis of academic achievement, a reading score for the verbal measure of student achievement is used.[18]

- ***Using Statistical Controls for Achievement and Dietary Intake Outcomes.*** Regression adjustments are included for the achievement and dietary outcomes. The analysis of achievement includes a measure of prior achievement; the analysis of dietary intake does not include a pre-measure of intake, since it has very weak explanatory power.

---

[16]For simplicity, all of the calculations presented in the report assume a one-tailed test. However, some of the hypotheses that will be tested will use two-tailed tests. For those outcomes, the minimum detectable differences will be higher than those shown in the tables, by approximately 12 percent.

[17]An unbalanced design, in which more schools or students are assigned to the control group, would save demonstration (breakfast subsidy) costs. For practical purposes, in the discussion of this chapter, unbalanced versions of the designs yield approximately the same magnitudes of precision shown in the text for balanced designs.

[18]These data are from a large, single school district in Minneapolis that used the California Achievement Test. Precision estimates were conducted using vocabulary as well as math computation and concepts measures. The results are similar to those reported in the text for reading achievement.

3. Factors Affecting Precision Levels

In order to understand the precision level findings presented later, it is useful to focus on two factors associated with the planned USBP evaluation which have important effects on attainable precision levels. The two factors are:

- **The "Dilution" Effect.** The USBP will likely affect only the students who actually start receiving school breakfasts under the USBP, and this will be a minority of students. Many students--on average, 25 to 30 percent--will already be participating in the regular SBP, and their outcomes are not likely to be substantially affected by the introduction of school breakfasts since USBP and SBP breakfasts will likely have the same nutritional quality. Among the other students, significant numbers will continue to not participate in school breakfast, either because their families choose to continue to provide breakfast at home or for other reasons. The precision calculations are based on an assumption that the increase in school breakfast participation will be approximately 25 percentage points (that is, an increase from 30 to 55 percent). In the experimental-versus-control comparisons, which are the essence of the full application of the experimental design, the effects of school breakfast on those 25 percent who are new participants are diluted by the 75 percent of students whose participation is not changed, making it difficult to detect USBP impacts.

- **Variation Between Schools.** The second factor that may limit the precision in the demonstration is that, even with pairing of schools on measured characteristics prior to randomization, there are likely to be significant unmeasured differences between the schools. This variation, because it is correlated with treatment status, tends to confound the analysis, making it difficult to disentangle school differences from the effects of the USBP.

The possibility mentioned earlier and discussed later in this chapter, of conducting a preimplementation survey at the schools to identify the students most likely to participate in the USBP, represents a possible approach to dealing with the "dilution" problem. In addition, using analysis approaches which are not based directly on the experimental design represents a different way of addressing this problem (see discussion in Chapter VII). In particular, it would take advantage in the analysis of the fact that the evaluation can observe *which students actually*

participate in the USBP, thus making it possible to sharpen the comparisons in the analysis, though at the cost of lowering some of the advantages of the pure experimental design.

For a given number of schools, adding more children per school to the design has only very limited effects in solving the problem arising from school-level variation. At the limit, even if the evaluation had data on *all* the children in a school, and hence could measure the school means perfectly, there would still be only a limited number of school observations with which to estimate treatment effects. Rather, increasing the number of schools in the USBP and control group is the best method of reducing the variability of the impact estimator. However, the ability of the evaluation to increase the samples of schools is constrained by the amount of funding available for the demonstration and evaluation.[19]

4. Detailed Precision Estimates of the Effects of the USBP on Student Outcomes

This section presents estimated minimum detectable effects for the two alternative designs and options. The sample sizes (in terms of schools and students) that underlie the precision estimates reflect approximately the level of resources available for the demonstration. Essentially all the designs included in this overview section can be considered feasible for the demonstration.

For each of the alternatives and options, the middle of Table V.1 displays *minimum detectable effects*, which are defined as the smallest demonstration effects the analysis could reliably detect, if

---

[19]Randomizing students *within* schools represents a possible way to get past this problem by essentially eliminating the correlation between children's treatment status and school differences. But this approach has the drawback that the treatment itself less realistically mimics what would take place in a national program. For example, it is not clear that a program offered to, say, half the students in a school would draw in the same number of students who would be drawn in by a full-school program. This would be a problem not only for examining participation effects, but for examining achievement effects as well, since one does not know that the students drawn into the program are representative of those who would be in a full-school program. In addition, implementing this experiment would be complicated and schools would probably not participate. For these reasons, a design including within-school randomization of students was rejected.

TABLE V.1

OVERVIEW OF THE STATISTICAL PRECISION OF THE ALTERNATIVE IMPACT EVALUATION DESIGNS

	Design Alternative 1: Baseline and Followup District-Administered Achievement Data Available		Design Alternative 2: Evaluation Contractor Must Conduct a Follow-up Achievement Test	
	Without Preimplementation Survey[a]	With Preimplementation Survey[a,c]	Without Preimplementation Survey[a]	With Preimplementation Survey[a,c]
Sample Size[b]	144 schools; 4,320 students	144 schools; 4,320 students	120 schools; 3,600 students	120 schools; 3,600 students
MDEs: Experimental-Based Methods				
Participation (in percentage points)	6.7	6.7	7.3	7.3
Dietary Intake (percentage of RDA)				
Food energy	6.0	3.5	6.5	3.9
Calcium	11.4	6.6	12.5	7.3
Academic Achievement (percent of standard deviation)	27.2	16.0	29.8	18.0
MDEs: Nonexperimental-Based Methods				
Dietary Intake (percentage of RDA)				
Food energy	1.9	1.9	2.1	2.1
Calcium	3.6	3.6	3.9	3.9
Academic Achievement (percent of standard deviation)	8.7	8.6	9.5	9.4

NOTES: See tables in Section V.C.4 for formulas used to calculate precision of estimates.

[a] If the demonstration startup is delayed to January 2001, then the evaluation will include a preimplementation survey to attempt to identify students in the treatment and control groups most likely to become new breakfast program participants. Precision estimates shown assume preimplementation survey predicts new participants at least 50 percent of the time (that is, of the parents who are surveyed that say their children would become new participants and being compared in the treatment and control groups, 50 percent of the students would, in fact, become new participants under the USBP).

[b] The design is a "balanced" design, in which half of the school s are USBP schools and half regular SBP schools. Demonstration costs will be lower if an "unbalanced" design, that is, one with more SBP than USBP schools. Statistical precision will be somewhat lower under an unbalanced design.

[c] Calculations assume that the preimplementation survey method will enable the evaluation to accurately predict "would be new participants" at least 50 percent of the time (that is, of the students currently not participating in the regular SBP who are surveyed that say they would participate in the USBP and being compared in the treatment and control groups, 50 percent of them would in fact become new participants under the USBP).

MDE = minimum detectable effect

indeed there are effects. These minimum detectable effects are shown for four illustrative outcome measures: school breakfast program participation, intake of food energy, intake of calcium, and scores on a standardized reading test.

a. Precision for Detecting Impacts on USBP Participation

It is estimated that between 25 and 30 percent of students attending *elementary schools* participating in the regular SBP receive a school breakfast on a typical school day (Burghardt et al. 1993a). Since there have been only a few evaluations of a universal-free school breakfast program, and none have used random assignment, it is difficult to know what kinds of effects on breakfast program participation to expect and plan for in the USBP demonstration. Based on existing literature and the informed opinions of experts, it is expected that participation in USBP schools would increase by 25 percentage points.[20]

As shown in Table V.2, both design alternatives would be able to detect the effects on participation considerably smaller than the effects expected for the demonstration. For example, under Design 1, with 144 schools (72 treatment and 72 control), an analysis of student participation

---

[20]Increases in student participation as low as 12 percentage points (an increase from 15 to 27 percent) have been found in a study of three schools in Baltimore and Philadelphia school districts in which the programs provided breakfasts in the cafeteria without any enhancements (Murphy et al. 1998). On the other hand, if schools make the program part of the school day (for example, by linking breakfasts to the classroom, either directly by serving breakfasts in the classroom, or more indirectly by having all members of the class go to the cafeteria as a group), then participation could increase by as much as 50 percentage points, where approximately 75 to 80 percent of students would participate in the USBP (Murphy et al. 1998b; Murphy et al. 1999; and CAREI 1997). Since SFAs will have to make application for the demonstration, it is likely that the schools would have a greater degree of enthusiasm about being selected; therefore, they might have a higher rate of participation than was observed in the Baltimore and Philadelphia school districts and what is typically observed for Provision II and III schools. The power calculations in this report assume an increase of 25 percentage points. This is consistent with the notion that most schools will offer the breakfasts in the cafeteria and operate the program essentially the same as the regular SBP, but with some enhancements to encourage participation.

TABLE V.2

MINIMUM DETECTABLE IMPACTS FOR THE UNIVERSAL-FREE SCHOOL BREAKFAST PROGRAM EVALUATION: PARTICIPATION EFFECTS
(Percent)

School/Students[a]	School-Level	Student-Level
120 Schools	8.3	
30 Students per School		7.3
144 Schools	7.6	
30 Students per School		6.7

SOURCE: Means, standard deviations, and cross-school and cross-food authority components of variance for the school breakfast participation were computed using data from the 1992 School Nutrition Dietary Assessment public use file. For documentation see Forkosh, Elyse, and Nancy Wemmerus. "School Nutrition Dietary Assessment Data File Documentation, Volumes I and II." Princeton, NJ: Mathematica Policy Research, February 1994.

NOTE: Minimum detectable effects are estimated for a simple one-tail test for differences of means between the treatment group and the control group, with a level of significance of 5 percent and a power level of 80 percent.

The formula for the school-level estimate is:

$$Detectable\ Effect = 2.49\sigma\sqrt{(\frac{1}{AS^T}) + (\frac{1}{AS^C})}$$

The formula for the student-level estimate is:

$$Detectable\ Effect = 2.49\ \sigma\sqrt{(1 - \theta - \gamma)(\frac{1}{AS^TN^T} + \frac{1}{AS^CN^C}) + \theta\,[(\frac{1}{AS^T}) + (\frac{1}{AS^C})]}.$$

where the first term under the square root is student-level variance and the second term is cross-school variance. In particular, σ is the standard deviation, θ is the proportion of variance due to cross-school variance, γ is the proportion of variance due to cross-food authority variance, N^T and N^C are the number of students sampled in each of the treatment and control schools, S^T and S^C are the number of treatment and control schools in each

TABLE V.2 *(continued)*

school food authority, and A is the number of school food authorities. The formulas are set up to assume that the results are being generalized only to the participating food authorities. For this reason, there is no cross-school food authority term in the formula.

[a]The computations assume that the number of observations in each group (treatments versus controls) being compared is half the sample size listed. For example, in the design with 120 schools, 60 are in the treatment group and 60 are in the control group.

based on school averages from administrative records data would be able to detect an increase in participation as low as 7.6 percentage points. If the analysis were conducted using microlevel data on individual students, the precision would be somewhat greater. For analyses at the student level, a 144-school design with 30 students sampled per school (a total of 4,320 students) would detect participation impacts as small as 6.7 percentage points.

Under the design that involves randomizing 120 schools, the evaluation could detect participation impacts as small as 8.3 percentage points using data from school administrative records. If the participation analysis is conducted using microlevel data on a sample of 3,600 individual students, the precision would be 7.3 percentage points.

b. Precision for Detecting Impacts on Dietary Intake

The minimum detectable differences in regression-adjusted mean differences in dietary intake, expressed as a percentage of the Recommended Dietary Allowance (RDA) at breakfast, for two nutrients, food energy and calcium, are presented in this section. In order to provide a reference point against which to consider the minimum detectable differences that are presented, note that previous research has estimated that eating a school breakfast increases students intakes of food energy relative to the RDA by 6 percentage points, based on data from SNDA (Burghardt et al. 1993a): it is estimated to increase *participating* student's intake of calcium relative to the RDA by 11 percentage points.

Detectable Impacts on Dietary Intake Achieved Without Conducting a Preimplementation Survey. Under Design 1, with 144 schools and 4,320 students, the evaluation would be able to detect USBP effects on food energy, expressed as a percentage of the RDA, of about 6.0 percentage points for the students affected (see Table V.3), using experimental-based analysis

TABLE V.3

MINIMUM DETECTABLE USBP IMPACTS ON DIETARY INTAKE WITHOUT PRE-IMPLEMENTATION SURVEY
(Student-Level Analysis: Comparisons of Means of the Full Treatment and Control Group Samples)

	Minimum Detectable Difference (Percentage Points) Intake Relative to the RDA	
Sample Size[a]	Food Energy	Calcium
120 schools/3,600 students	6.54	12.54
144 schools/4,320 students	5.96	11.44

SOURCE: Means, standard deviations, and cross-school and cross-food authority components of variance for the dietary measures were computed using data from the 1992 School Nutrition Dietary Assessment public use file. For documentation see Forkosh, Elyse and Nancy Wemmerus. "School Nutrition Dietary Assessment Data File Documentation, Volumes I and II." Princeton, NJ: Mathematica Policy Research, February 1994.

NOTE: Minimum detectable effects are estimated for a one-tailed test for differences of means between the treatment group and the control group, with a level of significance of 5 percent and a power level of 80 percent. Minimum detectable effects shown in the table equal the detectable effect (derived from the formula below), divided by the percentage of students in the USBP who would become new participants (that is, did not participate under the regular SBP but participate under the USBP), which is assumed to equal 25 percentage points.

$$Detectable\ \ Effect = 2.49\,\sigma\sqrt{(1-\theta-\gamma)(1-R_1^2)\left(\frac{1}{AS^TN^T}+\frac{1}{AS^CN^C}\right)+\theta(1-R_2^2)\left(\frac{1}{AS^T}+\frac{1}{AS^C}\right)}.$$

where the first term under the square root is student-level variance and the second term is cross-school variance. In particular, σ is the standard deviation, θ is the proportion of variance due to cross-school variance, γ is the proportion of variance due to cross-school food authority variance, N^T and N^C are the number of students sampled in each of the treatment and control schools, S^T and S^C are the number of treatment and control schools in each school food authority, R_1^2 is the proportion of cross-school variance explained by a regression model estimated on school means, R_2^2 is the proportion of within-school individual-level variance explained by a regression model estimated on the deviations of individual-level characteristics from school means, and A is the number of school food authorities. The formula is set up to assume that the results are being generalized only to the participating school food authorities. For this reason, there is no cross-food school authority term in the formula.

The table entries represent the minimum detectable impacts on student outcomes based on a comparison of outcomes for the full samples of students in the treatment and control schools, adjusting for the fact that only 25 percent of treatment group members will be new participants.

[a] Calculations assume that there are 30 students per school. It is also assumed that the schools participating in each food authority are split evenly between treatment and control schools. For example, in the design with 120 schools, 60 are in the treatment group and 60 are in the control group.

BEST COPY AVAILABLE

methods.[21] Thus, the evaluation would be able to reliably detect the benchmark effect defined above, of 6 percentage points. The 144 school/4,320 student design will be able to detect a program effect on calcium of 11.4 percentage points, somewhat higher than the benchmark (see Table V.3).[22]

Under Design 2, with 120 schools and 3,600 students, the evaluation would be able to detect USBP effects on food energy, expressed as a percentage of the RDA, of about 6.5 percentage points for the students affected, using experimental-based analysis methods.[23] This is 0.5 percentage points above the target benchmark. The 120 school/3,600 student design will be able to detect a program effect on calcium of 12.5 percentage points, which is 1.5 percentage points higher than the benchmark.[24]

Detectable Impacts on Dietary Intake Achieved Conducting a Preimplementation Survey. If the evaluation, on the basis of a baseline questionnaire, were able to predict relatively accurately

---

[21]With a sample of 144 schools and 4,320 students, the minimum detectable difference in food energy intake relative to the RDA equals 1.49 percentage points. However, the USBP will, for the most part, affect only students who are new participants, which is assumed to be 25 percent of students. The effect on the overall sample mean will be diluted by all the children (75 percent of the sample) who do not experience any effect. So, in order to detect an effect of 1.49 percentage points on the sample mean, the evaluation would need to observe a larger effect on the students affected. Under the 144 school/4,320 student design, the minimum detectable difference on newly participating students would need to be 5.96 percentage points (1.49 divided by .25), in order to detect a 1.49 increase in food energy intake relative to the RDA, when comparing the full treatment and control group samples. This is within the target of 6 percentage points.

[22]The minimum detectable difference in calcium intake relative to the RDA equals 2.86 percentage points. To detect an effect of 2.86 percentage points on the sample means, the minimum detectable difference on newly participating students would need to be 11.4 percentage points (2.86 divided by .25), for the full sample.

[23]Under the 120 school/3,600 student design, the minimum detectable difference is 1.63 percentage points. However, the evaluation would need to detect an impact on newly participating students for food energy of 6.5 percentage points (1.63 divided by .25), to detect a change of 1.63 percentage points in the full sample.

[24]The minimum detectable difference in calcium intake relative to the RDA equals 3.13 percentage points. To detect an effect of 3.13 percentage points on the sample means, the minimum detectable difference on newly participating students would need to be 12.5 percentage points (3.13 divided by .25), for the full sample.

which students (in both treatment and control schools) would participate because of the demonstration, the focus of the analysis could be on those students and thus improve statistical precision and power. Operationally, under this approach, the evaluation would administer a preimplementation survey to the parents of sampled students in each demonstration school to categorize them as either: (1) Group 1: current regular-SBP participants; (2) Group 2: current regular-SBP nonparticipants who would likely be new participants in a USBP; or (3) Group 3: current regular-SBP nonparticipants who would not likely participate in the USBP. The evaluation would then estimate impacts on students by comparing outcomes for Group 2 sample members in treatment and control schools. It would oversample Group 2 students somewhat (it is assumed here that 50 percent of the sample would come from Group 2).

The ability of this approach to detect program impacts on dietary intake, should they exist, depends crucially on the accuracy with which it predicts students who would become new USBP participants. Table V.4 illustrates the trade-off between statistical precision of impact estimates and the accuracy of the approach that identifies new participants, for the 144 school/4,320 student design.

If the method of predicting new participants is accurate 50 percent or more of the time, the evaluation would be able to detect the benchmark dietary impacts with the 144 school/4,320-student design (see Table V.4). For example, it would detect USBP effects on the intake of food energy relative to the RDA equal to 3.5 percentage points.[25] This is well below the SNDA breakfast

[25]If all demonstration school Group 2 members actually participated in the breakfast program, a design with 144 schools and 4,320 students will detect effects equal to 1.76 percentage points in food energy intake for the Group 2 sample, assuming Group 2 consists of 2,160 students (50 percent of the USBP and regular-SBP samples). However, the USBP will affect only students in Group 2 who are, in fact, new participants; the overall Group 2 mean will be diluted by students who do not participate and experience the effect. If the mechanism used to identify new participants (Group 2) is accurate 50 percent of the time, the minimum detectable difference on newly participating students would need to be 3.5 percentage points (1.76 divided by .5), in order to detect a 1.76 percentage point increase in food energy intake relative to the RDA, for the full Group 2 sample.

TABLE V.4

MINIMUM DETECTABLE DIFFERENCES IN DIETARY INTAKE UNDER DESIGN 1B
BASED ON COMPARISONS OF POTENTIAL PARTICIPANTS

Accuracy of Approach[a]	Number of Sampled Students per School[b]	Fraction of Student Sample Assigned to New Participant Group	Minimum Detectable Impact	
			Food Energy (as percent of RDA)	Calcium (as percent of RDA)
100 percent	30	.5	1.76	3.32
75 percent	30	.5	2.35	4.43
50 percent	30	.5	3.52	6.64
33 percent	30	.5	5.33	10.06

SOURCE: Means, standard deviations, and cross-school and cross-food authority components of variance for the dietary measures were computed using data from the 1992 School Nutrition Dietary Assessment public use file. For documentation see Forkosh, Elyse and Nancy Wemmerus. "School Nutrition Dietary Assessment Data File Documentation, Volumes I and II." Princeton, NJ: Mathematica Policy Research, February 1994.

NOTE: Minimum detectable effects are estimated for a one-tailed test for differences of means between the members of the treatment group and the control group students who would most likely become new participants in the USBP, with a level of significance of 5 percent and a power level of 80 percent. The minimum detectable effect equals the detectable effect calculated by the formula shown below, divided by the "accuracy rate" of the approach used to identify students who would participate in the USBP:

$$Detectable\ \ Effect = 2.49\,\sigma \sqrt{(1-\theta-\gamma)(1-R_1^2)\left(\frac{1}{AS^TN^T}+\frac{1}{AS^CN^C}\right)+\theta(1-R_2^2)\left(\frac{1}{AS^T}+\frac{1}{AS^C}\right)},$$

where the first term under the square root is student-level variance and the second term is cross-school variance. In particular, σ is the standard deviation, θ is the proportion of variance due to cross-school variance, γ is the proportion of variance due to cross-school food authority variance, N^T and N^C are the number of students sampled in each of the treatment and control schools, S^T and S^C are the number of treatment and control schools in each school food authority, R_1^2 is the proportion of cross-school variance explained by a regression model estimated on school means, R_2^2 is the proportion of within-school individual-level variance explained by a regression model estimated on the deviations of individual-level characteristics from school means, and A is the number of school food authorities. The formula is set up to assume that the results are being generalized only to the participating school food authorities. For this reason, there is no cross-food school authority term in the formula.

[a] The accuracy rate is the fraction of students who report in the preimplementation survey that they would participate in the USBP who, in fact, are new participants.

[b] It is assumed that the schools participating in each food authority are split evenly between treatment and control schools; overall, 144 schools are included (72 in the treatment group and 72 are in the control group). Thirty students are sampled from each school. Of this number, 15 are students who were not participating in the regular SBP but reported in the preimplementation survey that they would participate in a USBP.

BEST COPY AVAILABLE

program impact on food energy intake of 6 percentage points of the RDA. A similar conclusion holds when calcium is the outcome. The evaluation can detect a program impact of 6.6 percentage points in calcium intake relative to the RDA, which is substantially less than the SNDA estimate of 11 percentage points. Table V.4 also shows that if the mechanism for identifying would-be new participants in the Group 2 treatment and control samples were less accurate (correct only one-third of the time), this design could still reliably detect the benchmark program effects.

Similar conclusions hold under the 120 school/3,600 student design (tabulations not shown). Under Design Alternative 2, the evaluation would be able to detect impacts on food energy and calcium as low as 3.9 and 7.3 percentage points, respectively, when the preimplementation survey is accurate at least 50 percent of the time. The evaluation will just be able to detect the target impacts, if the mechanism is accurate one-third of the time.

c. Precision of Impact Estimates of USBP on Student Academic Achievement

No evaluation of the USBP or SBP has been undertaken that has used random assignment, and only a few nonexperimental analyses of either the USBP or regular SBP have examined impacts on student achievement. This makes it difficult to know what kinds of program effects to expect and plan for. However, based on a review of the literature and discussions with researchers in the education area, a benchmark effect size of .08 percentage points of a standard deviation has been set, which is equivalent to an average gain in nationally normed test scores of about 3 percentile points.[26]

---

[26]A school-based demonstration that assessed the impact of smaller class size in the earliest grades on student math achievement found impacts ranging from .17 to .25 standard deviations when comparing reduced-size classes (classes in the range of 13 to 17 students) and regular-size classes (in the range of 22 to 25 students) with or without teacher aides (Mosteller 1995). Put differently, gains in the range of .17 to .25 standard deviations in achievement means a student at the 50th percentile of all students would move up to between the 57th and 60th percentile. That represents a considerable gain in academic performance. It is expected an intervention that reduces class size this dramatically would have a substantially larger impact on student achievement than a program

(continued...)

This rest of this section describes the statistical precision and power of analyses of student achievement under the four design alternatives.[27]

Statistical Precision of Student Achievement Using Experimental-Based Analysis Methods. Estimated statistical precision levels based on experimental-based analysis methods are substantially lower for analysis of school achievement than for dietary intake. Under the 144-school, 4,320-student design, the evaluation could at best detect an impact on student's reading achievement equal to about 27 percentage points of a standard deviation (an effect size of .272); this represents a change in 11 percentiles for a student at the 50th percentile (see Table V.5).[28] It turns out the evaluation would require approximately 450 schools and 14,000 students to detect the benchmark effect using experimental-based methods. This clearly is not feasible.

If the evaluation were able to draw on a relatively accurate pre-implementation survey to focus the treatment-control group comparisons of means to members of the student sample predicted to be most likely to become new participants, then, under Design 1, it would be able to detect impacts on student achievement of 16 percentage points, or a change of 6 percentiles (see Table V.6).

---

[26](...continued)
offering students a free breakfast (when, prior to that, students could either obtain the breakfast free or purchase one at a fairly modest price). Thus, it is more likely that impacts will be substantially smaller in the USBP evaluation. Meyers et al. (1989), in a study of the SBP, found that the SBP increased academic achievement of low-income students by 0.10 of a standard deviation. Since that analysis focused on low-income children and the intervention was the regular SBP relative to a no-SBP environment, it probably represents an upper-bound estimate of the impacts that could be expected to occur in the USBP demonstration. Consequently, a somewhat lower benchmark, of 0.08 of a standard deviation, has been chosen.

[27]The calculations for student achievement underlying the discussion in this section assume a one-tailed test. If a two-tailed test is assumed, then the minimum detectable differences increase somewhat, by approximately 12 percent.

[28]The 144 school/4,320 student design could detect an impact of 6.8 percentage points of a standard deviation. But with only 25 percent of the USBP sample expected to become new participants, the program effect would need to be 27.2 percentage points (6.8 divided by .25) for the evaluation to detect an 6.8 percentage point change for the full sample of students.

TABLE V.5

MINIMUM DETECTABLE USBP IMPACTS ON STUDENT ACHIEVEMENT WITHOUT PREIMPLEMENTATION SURVEY
(Student-Level Analysis: Comparisons of Means of the Full Treatment and Control Group Samples)

	Minimum Detectable Difference (MDD)	
Number of Schools[a]	Effect Size[b,d]	Percentile[c,d]
120	29.8	12
144	27.2	11

SOURCE: California Achievement Test results for third-grade achievement in reading is used for verbal computations, for a large school district in Minneapolis, Minnesota.

NOTE: Minimum detectable effects are estimated for a one-tailed test for differences of means between the treatment group and the control group, with a level of significance of 5 percent and a power level of 80 percent. Minimum detectable effects shown in the table equal the detectable effect (derived from the formula below), divided by the percentage of students in the USBP who become new participants (that is, who did not participate under the regular SBP but would participate under the USBP), which is assumed to equal 25 percentage points.

$$Detectable\ \ Effect = 2.49\,\sigma\sqrt{(1-\theta-\gamma)(1-R_1^2)\left(\frac{1}{AS^TN^T}+\frac{1}{AS^CN^C}\right)+\theta(1-R_N^2)\left(\frac{1}{AS^T}+\frac{1}{AS^C}\right)},$$

where the first term under the square root is student-level variance and the second term is cross-school variance. In particular, σ is the standard deviation, θ is the proportion of variance due to cross-school variance, γ is the proportion of variance due to cross-school food authority variance, N^T and N^C are the number of students sampled in each of the treatment and control schools, S^T and S^C are the number of treatment and control schools in each school food authority, R_1^2 is the proportion of cross-school variance explained by a regression model estimated on school means, R_N^2 is the proportion of within-school individual-level variance explained by a regression model estimated on the deviations of individual-level characteristics from school means, and A is the number of school food authorities. The formula is set up to assume that the results are being generalized only to the participating school food authorities. For this reason, there is no cross-food school authority term in the formula.

[a] Calculations assume there are 30 students per school. It is also assumed that the schools participating in each food authority are split evenly between treatment and control schools. For example, in the design with 120 schools, 60 are in the treatment group and 60 are in the control group.

[b] The MDDs for achievement are expressed in two ways. The entries in this column represent the percentage of a standard deviation (referred to in the literature as "effect size").

[c] The entries in this column represent the effect in percentiles, evaluated for a student at the 50th percentile.

[d] The models used to estimate the effect on achievement tests controls for prior achievement and socioeconomic characteristics (SES), and includes the race and special education status of the student, the poverty status of the family, and whether the parent was a single parent.

BEST COPY AVAILABLE

TABLE V.6

MINIMUM DETECTABLE DIFFERENCES IN STUDENT ACHIEVEMENT UNDER DESIGN 1B BASED ON COMPARISONS OF POTENTIAL PARTICIPANTS

Accuracy of Approach[a]	Number of Sampled Students per School[b]	Fraction of Student Sample Assigned to New Participant Group	Minimum Detectable Impact on Reading Achievement	
			Effect Size[c,e]	Percentiles[d,e]
100 percent	30	.5	8.0	3
75 percent	30	.5	10.7	4
50 percent	30	.5	16.0	6
33 percent	30	.5	24.2	10

SOURCE: California Achievement Test results for third-grade achievement in reading for the verbal, based on test data from a large school district in Minneapolis, Minnesota.

NOTE: Minimum detectable effects are estimated for a one-tailed test for differences of means between the members of the treatment group and the control group students who would most likely become new participants in the USBP, with a level of significance of 5 percent and a power level of 80 percent. The minimum detectable effect equals the detectable effect calculated by the formula shown below, divided by the "accuracy rate" of the approach used to identify students who would participate in the USBP:

$$Detectable\ \ Effect = 2.49\,\sigma\sqrt{(1-\theta-\gamma)(1-R_1^2)\left(\frac{1}{AS^TN^T}+\frac{1}{AS^CN^C}\right)+\theta(1-R_2^2)\left(\frac{1}{AS^T}+\frac{1}{AS^C}\right)}.$$

where the first term under the square root is student-level variance and the second term is cross-school variance. In particular, σ is the standard deviation, θ is the proportion of variance due to cross-school variance, γ is the proportion of variance due to cross-school food authority variance, N^T and N^C are the number of students sampled in each of the treatment and control schools, S^T and S^C are the number of treatment and control schools in each school food authority, R_1^2 is the proportion of cross-school variance explained by a regression model estimated on school means, R_1^2 is the proportion of within-school individual-level variance explained by a regression model estimated on the deviations of individual-level characteristics from school means, and A is the number of school food authorities. The formula is set up to assume that the results are being generalized only to the participating school food authorities. For this reason, there is no cross-food school authority term in the formula.

[a] The accuracy rate is the fraction of students who report in the preimplementation survey that they would participate in the USBP that in fact are new participants.

[b] It is assumed that the schools participating in each food authority are split evenly between treatment and control schools; overall, 144 schools are included (72 in the treatment group and 72 are in the control group). Thirty students are sampled from each school. Of the 30 students, 15 are students who were not participating in the regular SBP but reported in the preimplementation survey that they would participate in a USBP.

[c] The MDDs for achievement are expressed in two ways. The entries in this column represent the percentage of a standard deviation (referred to in the literature as "effect size").

[d] The entries in this column represent the effect in percentiles, evaluated for a student at the 50th percentile.

[e] The models used to estimate the effect on achievement tests controls for prior achievement and socioeconomic characteristics (SES), and includes the race and special education status of the student, the poverty status of the family, and whether the parent was a single parent.

BEST COPY AVAILABLE

Although an improvement over the design without the pre-implementation survey, this impact is still much larger than what one could reasonably expect to occur under the USBP.

Statistical Precision of Nonexperimental Estimates of USBP Effects on Student Outcomes. Under all the evaluation design approaches, there is a fallback position based on direct multivariate modeling of the effects of participation (rather than on "availability"). Adopting an analysis approach that is not entirely experimentally based and that takes advantage of the fact that the evaluation can observe which students actually participate, the evaluation can model the effects of participation by comparing directly USBP participants and nonparticipants in the control group on student outcomes. (This line of analysis is described in detail in Chapter VII.)

How well this approach would work depends on the degree of selection bias in students' participation analysis and the degree of the evaluation contractor's success in correcting for such bias, should it exist. Selection bias is present if unmeasured characteristics of individuals affect their decision to participate and these characteristics affect outcomes. Techniques for correcting for this problem exist; however, it may be difficult to control for all the relevant factors.

If selection bias is not present, then, as summarized in Table V.1, the multivariate approach will allow the evaluation to detect impacts on student academic achievement quite close to the benchmark. Under Design 1, the evaluation could detect a change in student achievement of 8.7 percentage points of a standard deviation using nonexperimental methods; this is 0.7 percentage points above the target. Design 2, because of smaller sample of schools and students, could detect an impact that is somewhat higher--equal to 9.5 percentage points.[29]

---

[29]The formula for calculating precision estimates with nonexperimental methods is similar to that used for experimental-based methods (see bottom of Table V.5), except that sample sizes are different. For Design 1 (without a preimplementation survey), 540 new USBP participants from the treatment schools are being compared against 1,512 nonparticipants from SBP control group schools. (continued...)

Precision is only slightly improved if the evaluation includes a preimplementation survey and it is reasonably accurate. Under Design 1, the evaluation could detect a change in student achievement of 8.6 percentage points of a standard deviation using nonexperimental methods and preimplementation survey. Design 2 could detect an impact equal 9.4 percentage points.[30]

However, if selection bias is present, then methods that take it into account could potentially exacerbate the precision problem. This is the case, because statistical methods that control for selection bias are based on finding a variable that is a good predictor of participation but that does not affect the outcome. Finding a predictor with these qualities is usually difficult in practice, and almost always open to controversy.

A reasonable estimate is that, if selection bias is present, the multivariate methods available for dealing with it would result in estimated standard errors increasing by a factor of 3 to 4 times, requiring sample sizes nearly 10 times as large, assuming that the underlying model specification is correct. Consequently, if selection bias is important, the evaluation would be unable to reliably estimate impacts on dietary intake and achievement outcomes in the USBP demonstration.

---

[29](...continued)
For Design 2 (without a preimplementation survey), 450 new USBP participants are being compared with 1,260 nonparticipants from SBP schools.

[30]For Design 1 (with a preimplementation survey), 569 new USBP participants from the treatment schools are compared against 1,512 nonparticipants from SBP control group schools; for Design 2 (with a preimplementation survey), 474 new USBP participants are compared with 1,260 nonparticipants from SBP schools.

VI. DATA COLLECTION PLANS

A successful evaluation of the Universal-Free School Breakfast Program (USBP) demonstration will require the evaluation contractor to carefully collect accurate and timely information on a variety of student and school characteristics and outcomes. This chapter describes the sources of this information and discusses plans for how the evaluation will collect the data. It describes the data collection instruments and the modes by which the data are to be collected, identifies respondents, and provides information on when the data would be collected and where interviews would occur (for example, at school versus at students' homes). More specific details on data collection plans--such as how the data collection would be coordinated, the number of interviewers needed to collect the data, time spent at school districts, and so on--are not described in the report, since those plans would be specified by evaluation contractors during the proposal process.

Section A of this chapter briefly describes the different types of data collection instruments and methodologies that will be carried out as part of the impact evaluation. The major issues surrounding each type of data collection are discussed separately in Sections B through D for school records data, surveys of students and their parents, and surveys of teachers.

A. OVERVIEW OF DATA COLLECTION INSTRUMENTS/METHODS

The USBP evaluation will undertake a wide range of data collection activities involving the collection of school administrative data, the administration of cognitive tests (and possibly achievement tests), surveys of students and their parents, and site visits (see Table VI.1). The data will be collected from students, parents, and teachers and extracted from school records. Collecting data from many different sources, and using a variety of methods, will allow the evaluation to obtain accurate, reliable information on a wide range of school and student characteristics and outcomes.

TABLE VI.1

USBP DEMONSTRATION EVALUATION DATA COLLECTION INSTRUMENTS/METHODS AND DOMAINS FOR THE IMPACT STUDY

Data Collection Instrument/Method	Domain									
	USBP/SBP Participation	Breakfast Consumption	Dietary Intake	Food Security	Academic Achievement	Attendance/ Tardiness	Disciplinary Behavior	Cognitive Function Behavior	Health	School Environment
Administrative Data										
School Records					✔	✔	✔		✔	
School/Food Service Personnel										
Cafeteria Manager Survey	✔									
School Administrator Survey							✔			✔
Teacher Survey							✔	✔		✔
Students and Parents										
Preimplementation Parent Survey*	✔	✔								
Parent Survey	✔	✔		✔		✔			✔	
Student Survey	✔	✔								✔
24-Hour Dietary Intake	✔	✔	✔							
Height/Weight									✔	
Achievement Test Administration					✔					
Cognitive Performance Test Administration								✔		
Observation of School Meals	✔									

*The preimplementation survey will collect information on student's current SBP participation, intentions to participate in the USBP, parental attitudes toward breakfast and school breakfast, and household and student characteristics

The discussion below briefly describes the school records available for the students and the school breakfast program, student and parent data collection, and data that would be collected from teachers as part of the evaluation.

Ideally, the school districts selected for the USBP demonstration will maintain administrative data that are well documented and easily extracted. The administrative data will serve as a key source of information on school characteristics and both school-level and student-level outcomes, especially for academic achievement. For example, it is expected that demonstration schools will be able to provide administrative data on participation in the school breakfast program, academic outcomes such as test scores, behavioral outcomes like school attendance, and school characteristics such as total enrollment and the racial composition of the school.

Data on school breakfast participation will be collected through several sources: administrative data (school records), surveys of parents and students, and observation of school breakfast participation during a target week. The 24-hour dietary recall with the student will provide information on USBP/SPBP participation on the target day. Students and parents will report usual participation.

Surveys of school personnel will be used to provide information on school outcomes, as well as student outcomes. General school characteristics, as well as data on the length of the school day, will be collected as part of the cafeteria manager's survey. The teacher's survey will provide information on students' usual behavior and cognition.

Along with district-provided standardized test data on student achievement, a critical source of information for the evaluation will be data collected from students and their parents. The student survey will provide information on student characteristics, attitudes about school breakfast, and student outcomes. The parent survey will provide additional information on student characteristics,

family characteristics such as income and food security, and parental satisfaction with the breakfast program in their child's school. Dietary intake data will be collected from students and parents of young students. This information will be collected using a 24-hour dietary recall survey; a second recall, administered approximately within one week to 10 days of the first, will be administered to a subsample of students for the purposes of estimating usual dietary intake for the entire student sample. Additional information will be obtained from students through the administration of short-term cognitive tests to measure their memory and cognitive functioning (and, if necessary, standardized tests to measure students' academic achievement).

B. SCHOOL RECORDS DATA COLLECTION

Examining the degree to which the USBP improves school outcomes for participating children is a central component of the planned evaluation. Through the collection of school records data, the evaluation will obtain accurate and timely data on a variety of student and school outcomes with minimal respondent burden. Schools' administrative records are potentially a rich source for both school-level and student-level data on program participation, attendance, tardiness, test scores, visits to the school nurse, height and weight, and behavioral characteristics such as disciplinary incidents. Other school records will be used by the evaluation, including student rosters to sample students and teacher and class lists to identify teachers of the sampled students. Administrative school records are also the source for student rosters to be used for sampling and to identify teachers of the sampled students. In addition, administrative records contain background characteristics such as size of enrollment, ethnic composition, and other information of interest to the evaluation.

Most schools maintain all or many of the types of school-level outcomes data needed for the evaluation. For example, average daily attendance is kept by all schools to meet state and local education agency reporting requirements. School districts also require systematic reporting of test

scores and disciplinary actions. However, these records can vary across school districts. For example, the selection of specific standardized tests and the grades at which they are administered differ across school districts. Similarly, disciplinary records are not standardized nationally. Some districts require only that schools report expulsions or out-of-school suspensions lasting five days or more. Although individual schools are likely to keep records of their most serious disciplinary actions, children of elementary school age typically are unlikely to misbehave in ways that warrant these more serious actions. However, many schools also maintain records for minor disciplinary referrals to a "quiet room" or the school office, and many of them use suspensions both in school and out of school, from one to three days, as disciplinary actions. Since both record keeping and policies vary widely, some effort will have to be made to provide definitions of those types of actions that are consistent, as well as, perhaps, requesting that schools keep records of these actions. This may result in new work for some or all the schools in the demonstrations. However, the issue of student behavior is so important, both to program design and to educators in general, that it is important for the USBP evaluation to make the effort to define, standardize, and collect data on the varying levels of student disciplinary problems.[1]

Data on individual students can also be obtained from school records. Access to student rosters will be needed for the selection of the sample. During the evaluation, academic records--including test scores, attendance, visits to the nurse, and disciplinary action--are potentially available for individual members of the student sample.

---

[1]The limitations that will likely prevail in existing administrative data sets with regard to student disciplinary incidents and the low frequency of serious incidents in elementary school point to the need to supplement the administrative data with surveys of teachers, administrators, and other staff who would be able to provide better-quality data on student behavior. A number of teacher rating scales are available and have been used in previous studies of the school breakfast program; these scales should be added to the teacher questionnaires described to supplement the data on disciplinary incidents provided by administrative records.

1. Methods

The collection of administrative school records will be used to obtain many of the school outcomes data needed for the evaluation. The selection of data collection methodologies should take into consideration the best alternatives to the numerous challenges posed by the collection of individual student records from schools, as well as ways to collect comparable and accurate school-level data with minimal administrative burden. Two possibilities for collecting administrative records data are (1) to have schools complete an administrative data form for the school and for each sampled student, and (2) to have schools submit extant school records via electronic files.

Under the first method, and one to be used in the evaluation, an administrative data form designed by the evaluation team will be mailed to the school to collect data on the test scores, attendance and tardiness, disciplinary incidents, periods of enrollment, and other information for the school as a whole and for each student in the sample. The form would be designed with the dual objectives of collecting comprehensive and consistent data on student outcomes and being reasonably easy for school personnel to complete for a given student. A form would be printed out for the school and for each sampled student, then each school would be sent the appropriate set of forms.

The design and implementation of an administrative data form would ensure that the data arrive in consistent form. It would also encourage comparability of reporting, since items such as "disciplinary actions" or "tardiness" can be defined for the responding school. However, this method may increase burden for schools, especially those that do not define or store data in a way that is consistent with the form. This might lead to a decrease in accuracy (providing best estimates rather than exact numbers) or cooperation rates. Should schools have difficulty completing the form, the evaluation team should be prepared to offer technical assistance to help schools complete it.

A backup data collection method, to be used in the minority of schools that might refuse to complete forms, would be to collect school records in an electronic format, with data on individual students and/or the school as a whole (see discussion in Section VI.B.3.b). This would reduce the data collection burden on the school, but does involve some effort, such as extracting the appropriate outcome measures or suppressing data for nonsampled students. In addition, the data may not be complete or consistent across schools, or come with adequate documentation. After the evaluation team receives the electronic files, there may be an additional data processing step of making the information consistent across schools and districts.

2. Timing of Data Collection

The collection of both school-level and student-level administrative records will be ongoing throughout the evaluation, as shown in Table VI.2. School breakfast participation and school outcomes (attendance, tardiness, achievement test scores, disciplinary incidents, nurse visits, and height and weight) would be obtained each school semester beginning in fall 2000 and continuing through spring 2003. Data for fall 1999 and spring 2000 can be reported retrospectively, if available, and test scores can be obtained for spring 1999. If height and weight are available from school records, or the nurse is willing to maintain these records, they can be collected on the same schedule as other school outcomes. However, if the evaluation team takes these measurements, they would plan to measure sampled children in spring 2001 when obtaining dietary intake information. The disadvantage to the latter approach is that only comparisons between the USBP and SBP groups can be made, whereas change can be measured if the data for individual students are available over time.

TABLE VI.2

USBP EVALUATION: ILLUSTRATIVE DATA COLLECTION PLANS FOR THE IMPACT STUDY

Data Collection Plans	School Year 1999 - 2000		School Year 2000 - 2001		School Year 2001 - 2002		School Year 2002 - 2003	
	Fall 1999	Spring 2000	Fall 2000	Spring 2001	Fall 2001	Spring 2002	Fall 2002	Spring 2003
School Administrative Records (School-Level)								
School Breakfast Program participation	x	x	x	x	x	x	x	x
Achievement tests (district-administered test data)		x		x		x		x
Student behavior (attendance, disciplinary actions, health)	x	x	x	x	x	x	x	x
Administrative records data for *sampled students*	x	x	x	x	x	x	x	x
Teacher Survey				x				
Student/Parent Survey								
Student/family characteristics			x[a]	x				
School Breakfast Program participation			x[a]	x				
Dietary intake				x				
Height and weight measures								
Short-Term (Same-Day) Cognitive and Behavioral Testing				x				
Follow-up Academic Achievement Test				x[b]				

[a]Preimplementation survey to identify students in the treatment and control groups most likely to be new participants.

[b]This achievement test would be administered by the evaluation contractor, and would be conducted only if follow-up student achievement test data cannot be provided by participating school districts.

3. Issues

The methodologies used should anticipate issues surrounding parental consent for the release of student-level data and school cooperation issues.

a. Consent

Parental consent is necessary to obtain school records for individual students sampled. The evaluation team will negotiate with schools to determine what form of consent is needed, but it is likely that active consent will be required. That is, schools will require the evaluation team to produce a consent form signed by the parent or legal guardian for each child for whom individual school records will be collected. The consent form should be inclusive, with permission for the child to participate in all study activities, and it should specify the individual records on discipline, test scores, health, and attendance to be collected. It should also specify permission to provide free or reduced-price lunch or breakfast eligibility since, historically, schools and SFAs have been reluctant to provide this information for individual students. A signed parental consent form will be particularly important for records collected at subsequent points during the study, since attrition among school or food service staff or changes in school or district policies over time could render unacceptable any prior agreement to work with a passive-consent arrangement.

Regarding consent to obtain access to school records data, the data collection plan would be flexible to adopt whatever method of consent (either active or passive) that the district requires.

b. Submission of Existing or Electronic Files or Incomplete School Records

While the method to be used to collect consistent and comparable school outcome data will have schools complete an administrative records form developed by the evaluation team, it is possible that some schools will submit existing records rather than complete the requested form. The evaluation team will have to be prepared to work with the data provided and make determinations regarding its

quality and consistency with the study requirements. It is possible that the evaluation team, where possible, will need to be willing to work from these sources and to negotiate the completion of the school records form only for those data elements not obtainable from the school's existing format. Time and resources should be set aside for reconciliation of school records provided this way. Similarly, some schools might prefer to submit data electronically. The evaluation team, when possible, will encourage the school to incorporate the requirements of the study into their electronic reporting system and, with adequate documentation, simply capture needed data from the files. However, if the electronic files do not follow the study protocol, negotiations and followup to obtain necessary data elements must be conducted.

c. Summary

The benefits of receiving school records data in a consistent manner outweigh the costs of increasing the burden on schools of providing this information. Given that the number of students in any given school in the sample is likely to be relatively small (30 students), it is recommended that the evaluation use administrative data forms to collect the school records data on sampled students. Schools also would be given a form for entering schoolwide totals and averages (assuming either that they have already calculated these averages or that they themselves could easily do so). However, more contractor involvement may be necessary for some outcomes in which the school has not already calculated the school-level averages (or does not want to). In those cases, the contractor should be willing to work with school files for all students, stripped of all identifying information.

Obtaining good disciplinary data for young children from existing school records may not be feasible. These data will be supplemented, at least for individual student-level data collections, with behavior data obtained from teachers and, perhaps, more general information on school climate, classroom disruptions, and other characteristics obtained from school administrators or teachers.

using standardized behavior rating scales. Similarly, administrative data on student attendance or tardiness may not be adequate for purposes of the evaluation and will be supplemented with information reported by classroom teachers.

C. DATA COLLECTION FROM STUDENTS AND PARENTS

The USBP evaluation design needs to select appropriate methodologies and develop strategies for collecting demographic and socioeconomic information from parents, setting up and conducting dietary interviews with students and their parents, and conducting cognitive tests (and potentially administering achievement tests) to students. In addition, the evaluation may include a preimplementation survey of students' parents, in order to identify students most likely to become new USBP participants. The main considerations in the evaluation design should be cost and the ability to meet the study objectives of accurately measuring usual dietary intake and dietary intake on a target day, gathering sufficiently detailed information from students and their parents, and obtaining reliable measures of student cognitive functioning. The methods, timing of data collection, and issues for collecting information on student characteristics, dietary intake, achievement, and cognitive functioning are described below (see Table VI.3).

1. Dietary Intake

A combination of survey approaches should work best to estimate dietary intake and collect information about breakfast habits and the related variables needed to interpret dietary and nutrition findings. The dietary intake methodology for the evaluation study will include the collection of 24-hour dietary interviews with students and their parents to estimate quantitative food and nutrient intake data. The 24-hour recall data collection will characterize breakfast eaten at home and at school, and enable the estimation of nutrients wasted from school breakfast. Information on the

TABLE VI.3

OVERVIEW OF DATA COLLECTION INSTRUMENTS FOR THE USBP IMPACT EVALUATION

Instrument	Mode	Respondent	Timing of Data Collection
School Records Form	hard copy	school staff	ongoing--every semester fall 1999 - spring 2003
Observation of Student Participation (Target Week)	hard copy	n.a	spring 2001
Teacher Survey	hard copy mail survey[a]	teacher	spring 2001
Preimplementation Parent Survey[b]	telephone	parent	fall 2000
Student Parent Characteristics Survey	In-person at school, in-person or telephone followup at home	student parent	spring 2001
Student Dietary Recall	In-person at school, in-person or telephone followup at home	student parent	spring 2001
Cognitive tests	in-person at school	student	spring 2001
Achievement tests[c]	in-person at school or offsite location	student	spring 2001

[a] Forms will be mailed to teachers prior to school visits by field interviewers. Teachers will return completed surveys to field interviewers or mail them back.

[b] Preimplementation survey will only be conducted if demonstration startup is delayed until January 2001.

[c] A follow-up achievement test will be conducted by the research team only if follow-up standardized achievement test data cannot be provided by participating school districts.

student's vitamin/mineral supplement usage and food security (as well as other household members' food security) will be collected in the parent survey and used to evaluate overall dietary intake.

As discussed below, the need for high quality-dietary intake demands the use of a common dietary protocol for children in grades 1 through 5 or 6 (approximately age 6 to 11 years). The dietary reporting method will be a combination of self- and proxy-reporting where children report breakfast intake immediately following breakfast at school, and the child and parent report the rest of the day's intake together at a later interview the following day. In addition, the 24-hour reference period will be midnight to midnight on the target day; a second day of intake, using the same in-person methodology as the first day, will be collected on a representative 10-15 percent subsample of students to allow for the estimation of usual intake and adequacy of nutrient intake in USBP/SBP participants and nonparticipants.

a. Methods

The methodology of choice to produce quantitative food and nutrient intakes for children, and the one to be used in the USBP evaluation, is the 24-hour dietary recall. Some key data collection decisions to be made are: (1) whether to conduct multiple days of 24-hour recalls; (2) when and where to do the interview(s); (3) whether to use telephone versus in-person interviewing; and (4) how to incorporate parental assistance for child interviews, if needed. In conducting the 24-hour dietary recall interviews, parental consent and cooperation will be needed. In addition, the design needs to select appropriate dietary data collection protocols, as well as food composition databases and software for food and nutrient coding.

The dietary recall instrument. Several alternative versions of 24-hour recall instruments are available for use in the study. Possibilities include the instrument used by the most recent CSFII data collection, that used by NHANES 1999, and the one used for the SNDA study. Developmental work

is currently underway by the Agricultural Research Service of the USDA on a new instrument based on computerized interviewing methods. This method will be used in the new consolidated version of CSFII/NHANES, perhaps as early as 2001. Each of these instruments collects information on the type and amount of food eaten, along with the time the food was eaten, the meal name, and where it was obtained.

The 24-hour protocols used in CSFII and NHANES include detailed probes to elicit complete food intake, as well as probes for missing or frequently forgotten foods at the end of the recall. Standardized probes in the protocol serve to minimize problems with under-reporting of energy intake, a known systematic bias with 24-hour-recall methodology. However, energy under-reporting has been less well studied in young children than in adolescents and adults, and is likely less of an issue for interpreting the intakes of young children. The collection of measured height and weight of children would provide data to evaluate the extent of energy under-reporting and to interpret the completeness of the energy intake estimates in the study.

It is recommended that the evaluation use the most recent CSFII protocol. The 24-hour methodology used has been thoroughly tested through its use in the CSFII data collection, and the resulting data have proven extremely useful in examining USDA research issues through numerous studies.[2]

Multiple Observations of Dietary Intake. Collecting a single day of dietary intake data will allow us to measure students' mean intake but not the distribution of usual nutrient intake. Thus, the evaluation would not be able to estimate how USBP availability is related to the percentage of students whose usual intake of a given nutrient is above or below particular thresholds. The

---

[2]If the timing proves right, consideration should be given to the data collection and coding system currently under development by USDA for the joint CSFII/NHANES, scheduled to be merged in 2001.

evaluation can, however, estimate the distribution of usual intake if it collects data on a second day for at least a subset of the full sample. If two or more days of dietary intake data are collected, intakes on nonconsecutive days are preferable to intakes on consecutive days, since food choices on consecutive days may be correlated.

Using the same in-person methodology as the first day, the evaluation will collect a second day of intake data on a representative 10-15 percent subsample of students to allow for the estimation of usual intake and adequacy of nutrient intake in USBP/SBP participants and nonparticipants.

Nutrients Wasted. School breakfast menus and observations of school meals will provide information about foods and nutrients "available" for consumption by students. (The meals-offered methodology for evaluating foods served is fully described in the data collection chapter of the Implementation Study Design [Chapter VIII, Section B]). To estimate the nutrients wasted from school breakfast, the 24-hour-recall protocol will be modified to collect school breakfast information on foods taken and consumed. For each breakfast item selected, the student will be asked the portion or fraction of the food actually consumed: all, most (3/4), about half (1/2), almost none or just a bit (1/4), and none (0). This methodology was successfully used in SNDA-1 to calculate the percent of food wasted, foods selected for school breakfast but not consumed, and the nutrients wasted from the school breakfast. Selection of foods from school breakfast will be used to define school breakfast participation status on the target day.

b. Timing of Data Collection

If the demonstration is implemented in fall 2000, then dietary intake data would be collected once in spring 2001 in concert with same-day cognitive testing and the other student information, as shown in Table VI.2. A practical issue to consider when developing the dietary recall instrument are the start and ending times of the 24-hour recall period. Standard 24-hour dietary intake

interviews ask about the previous day's intake from midnight to midnight. If this approach were used in the USBP study, then interviews in the school to minimize recall error would not be useful because they do not correspond to the relevant breakfast period. An alternative approach would be to adapt the instrument to cover a period from noon on the previous day to noon on the current day. This has the obvious drawback of selecting a time that could interrupt an eating occasion (lunch) for many sample members. An alternative time, such as 10:00 A.M., could be used instead, or students could be asked about the 24-hour period preceding the interview, as was done in SNDA-1 (Burghardt et al. 1993b).

A preferred alternative to recalling foods over the previous 24 hours, and the one to be implemented in the USBP evaluation, is to split the 24-hour data collection by asking about all foods consumed since midnight the night before (or since waking up this morning), that is, breakfast consumption at home and/or at school, and collecting the rest of the target day's intake later that same day or the next day. Since a primary data need is to accurately assess breakfast intake on a target day, recall error can be minimized by interviewing children soon after breakfast, which would have to be done at school. Data collection at school has the advantage of allowing multiple interviews in a single location, and would likely produce the most accurate reporting of breakfast. A follow-up interview with the child and/or the child's parent or guardian would be needed to complete the day's intake for each child. Data collection of a second day's intake on a school day would occur approximately 7 to 10 days after the first day.

Finally, the collection of weekend days was considered to characterize children's usual intake, since breakfast patterns differ between weekdays and weekends. However, limiting the 24-hour-recall interviews to Mondays through Fridays would offer additional days of observation of breakfast consumption on school days. This would be preferable because school breakfast is the primary focus of this study, and because national data on children's weekend breakfast consumption are available

for comparison. It is important to collect data across all weekdays, thus requiring Saturday interviews to capture all of Friday's intake.

c. Issues

Although data collection at school may produce the most accurate picture of breakfast intake, it also has several disadvantages: (1) interruption of the school day, (2) interruption of the flow of the 24-hour-recall protocol, (3) possible altering of intake for the remainder of the day if the child or parent knows that intake information will be collected, and (4) a possibly longer 24-hour dietary interview by asking about a single day's intake at two different points in time.

Parental Assistance. The extent to which parents help their children in the dietary intake interviews is an important design issue. Adults often have difficulty recalling the timing and amounts of foods eaten; children may have greater difficulty. Some children at the elementary school level, particularly those age five to nine, may not have an adequate vocabulary or clear enough concepts of time and quantity to report accurately on what they have eaten, the amounts they have eaten, or the preparation method. On the other hand, schoolchildren may feel uncomfortable reporting actual food consumption at school with a parent present, especially if they did not follow the parent's wishes.

The 1994-1996 Continuing Survey of Food Intakes by Individuals (CSFII) collected dietary intake data from children age 6 to 11 years at home, with a parent or guardian present to assist the child. A similar protocol was used for dietary data collected in mobile examination centers in the Third National Health and Nutrition Examination Survey (NHANES III) in 1988-1994 and in the 1999 NHANES. Other caretakers for the child (such as day care providers) were consulted as needed if the child had difficulty reporting on intake away from home or at school. The SNDA study collected data differently for first and second graders than for older children (Burghardt et al. 1993b).

For first and second graders, the child was interviewed in school about foods eaten in school on the interview day, then interviewed later in the day at home, with a parent present, about foods eaten over the remainder of the 24-hour period (which started at the time of the interview on the previous day). Older children were interviewed in school about all foods they had eaten in the past 24 hours, starting with everything they ate since waking up that day, then going back to what they ate the previous day, from the time of the interview onward. Parents did not assist in the SNDA data collection for children older than second grade.

Timing and Location of the Dietary Recalls. At least three alternative approaches to collecting 24-hour recalls were considered when making the final decision for the evaluation study's design:

1. ***Conduct both in-school and at-home interviews on the same day.*** This would incur the greatest cost and burden, but likely would provide the most accurate responses. However, even this strategy could potentially fail to reach parents in the same evening as the target day for data collected at the school, and be impractical for interviewing parents late in the evening to obtain a full day's intake.

2. ***Conduct in-school interviews with students and make follow-up telephone calls to parents at the end of the target day or during the subsequent day*** to obtain intake data for a complete 24 hours.

3. ***Interview older children (or all children) without parental assistance.*** It is unlikely that young children can accurately report complete intakes or that teachers could accurately or easily provide information about individual students' breakfast intakes.

The timing of the interview, and who the respondent is, also has major implications for where the interviews take place. Dietary interviews conducted with parents or guardians assisting, as was done with children in past CSFIIs and NHANES, would require interviews at home (either in the evening on the target day or subsequent day) or follow-up telephone interviews. Ideally, the 24-hour recall would be completed as soon as possible after the child has eaten his or her last meal or snack on the target day. However, this would impose difficulties of collecting interviews late in the

evening of the target day. A more practical approach would be to collect the rest of the 24-hour recall the next day. The major disadvantage would be a loss of data if interviews could not be completed the next day. For interviews that had to be completed more than a day later, there may be increased recall error. Both in-home and telephone contacts would have drawbacks that interviewers would not necessarily succeed in completing interviews on the same day, or day after, they collected data in the school. This could undermine analyses that depended on same-day outcomes and total daily intake.

Interviewing at home allows parental assistance, but also has some advantages and disadvantages. The major advantages are, it (1) minimizes the burden on schools, (2) provides an opportunity to collect other interview data such as family background information at the same time, and (3) provides a contact and framework to facilitate follow-up dietary interviews for the collection of additional days of intake for the estimation of usual intake. The in-home option, however, may be more burdensome on the students' families than a telephone contact, and more costly, since it requires interviewers to visit the home of each sample member, rather than only each school.

The approach regarding the reporting method will be a combination of self- and proxy-reporting where children report breakfast intake immediately following breakfast at school, and the child and parent report the rest of the day's intake together at a later interview the following day.

Mode for Collecting Second Day Intake. When collecting two days of dietary intake, there may be problematic issues with mixing modes--that is, collecting the first day's intake by in-person interview and the second day by telephone. Little or no research is available on using the mixed-mode approach for children's intake data. First, there are known mode and sequence effects for 24-hour dietary recall data collected in person and by telephone. These may vary across nutrients and across individuals. Second, response rates may vary by mode and may potentially be lower for telephone interviews. Third, it may be difficult to adequately standardize portion size measures for

in-person and telephone data collection to allow for the data to be combined. Fourth, combining intake days collected by different modes for use in the Iowa State software to adjust nutrient distributions has not been well studied or resolved, especially for children's dietary intake data. Therefore, for the evaluation, both first- and second-day intakes will be collected using the same mode to minimize bias related to mode effects.

Food and Nutrient Database. Finally, choosing a food and nutrient database and associated software for coding the dietary intake data is another important issue to consider in the design. Some considerations in deciding on the database and associated coding software are:

- Which one is most current
- Which is better suited to assessing the dietary outcomes of interest (foods and nutrients)
- Which provides data that are most comparable to previous studies
- The cost and flexibility of the system

Either the University of Texas and the University of Minnesota 24-hour-recall dietary data entry and coding systems could be used for the USBP evaluation.

2. Student/Parent Surveys

Interviews with students and/or parents will be needed to obtain demographic and socioeconomic information and to fully assess usual breakfast patterns, attitudes about school breakfast, costs of breakfast, and breakfast program participation. In addition, both parents and children will be asked about their attitudes toward eating breakfast and school breakfast. These interviews would be used to collect information on children's use of dietary supplements and school

attendance, and on household food security. The evaluation may include a preimplementation survey in order to identify students who would be most likely to become participants in a USBP.

a. Methods

Preimplementation Survey of Parents. If demonstration startup is delayed, then the evaluation will conduct a survey of parents prior to program implementation. This survey would be used to identify students, in both treatment and control schools, most likely to become new participants under a universal-free program. These students would be oversampled and used in the impact analysis.

For the preimplementation survey, approximately 150 students would be sampled from rosters of enrolled students provided by the participating schools (see Table VI.3). The rosters would need to contain the names, addresses, and telephone numbers of students attending sampled treatment and control schools. A letter will be sent to the sampled student's home prior to the survey, describing the purpose of the survey; it will include an endorsement of the study by the school district. The student's parent will be surveyed by telephone; the interview would last 10 to 15 minutes. The survey would obtain information on children's participation in the regular SBP, the likelihood that the student will participate in a universal-free program, and attitudes about breakfast and the school breakfast program. The survey would also collect information on the household's socioeconomic characteristics and student characteristics.

Follow-up Surveys of Students and Parents. Students will be interviewed in school to collect information on school breakfast participation, usual breakfast patterns, attitudes toward school breakfast, and the school environment. Parents will be interviewed at home following the completion of the 24-hour dietary recall. They will provide information that will include household income, composition, food program participation, food security, attitudes toward school breakfast,

and potentially the student's attendance and academic achievement. Food security will be assessed using the 18-item household scale developed and first used in the 1995 Current Population Survey. These questions would be part of the parent survey rather than the student survey.

b. Timing of Data Collection

The preimplementation survey of students' parents would be conducted in fall 2000. The main (follow-up) student and parent interviews will be conducted in conjunction with the dietary recall interviews. For example, the parent interview will be conducted following completion of the 24-hour dietary recall with the student and parent. Given the burden of the dietary interviews, every effort should be made to limit the length of the student and parent interviews. Shorter interviews may limit the ability of the evaluation team to collect some of the information described above.

Table VI.2 illustrates the timing of data collection for one possible design for the evaluation study. Interviews with students and parents would occur in spring 2001, to provide information on the student's school breakfast participation and select family characteristics. Similar parent and student interviews to collect information on family and student characteristics and USBP/SBP participation would be conducted in spring 2002.[3]

c. Data Collection Issues

Although the food security questions may be considered sensitive by some respondents, these questions have been successfully administered in national surveys using in-person and telephone modes. If cost and respondent burden become a concern, use of the six-item food security short scale should be considered (Blumberg et al. 1999).

---

[3]If the preimplementation survey is conducted, data on school breakfast participation, attitudes toward breakfast, and student and family characteristics will be collected in fall 2000, prior to USBP implementation.

Based on NHANES III data, approximately 35 percent of U.S. children age 6 to 11 reported consuming at least one dietary supplement during the period studied (Ervin et al. 1999). Of particular interest for this study is collecting data on intake of supplements containing iron, folic acid, and B vitamins. Dietary supplement questions would be adapted from those used in the NHANES and CSFII studies, using the NHANES composition database for supplements.

It might also be desirable to ask parents or students about students' typical food consumption patterns for breakfast, using a food frequency type of interview. Food frequency-type information could provide qualitative information on the frequency and type of foods consumed for breakfast eaten at home, school, or some other location. Collection of this type of data, however, would increase respondent burden for the parent or student interview, and would not provide quantitative intake of foods or nutrients unless specific questions were asked about portion size, frequency, and quantity of the foods typically consumed for breakfast. Although potentially this is interesting data, it is not recommended that food frequency data be collected because of the burden to respondents.

3. Achievement Tests

Under the preferred design option, achievement test data would be available from school records. If follow-up achievement test data are not available from participating districts, the evaluation team would have to administer special achievement tests. If the evaluation contractor must administer an achievement test in the USBP evaluation, then it is recommended that the Iowa Test be used. However, as mentioned in Chapter IV, it may be prudent to postpone a final decision until it is determined what testing packages are currently in use in the school districts selected for the evaluation. If one of the above tests is used predominantly in those districts, then it may be appropriate to choose that package for the follow-up test, in order to allow greater comparability across sites.

The advantages and disadvantages of these two approaches, the criteria for the types of tests selected, and related issues are discussed in Chapter IV. The discussion here considers the logistics of implementing achievement testing as a data collection activity.

a. Method

A major logistical question is when and where to test students. Testing students in school during the school day would be ideal if cooperation could be obtained from parents and school staff. The room would be familiar to the students, teachers could serve as proctors, and the students would already be in school. The major difficulty would be consent of schools and teachers. Pulling students out of class for achievement testing is considered by many teachers to be burdensome and unproductive, interrupting the flow of instruction. This is particularly true if the test is a norm-referenced standardized test, which teachers often perceive as threatening. In some cases, limits on the amount of time devoted to standardized testing are specified in a teacher contract.

A standardized achievement test battery, such as the Iowa Test of Basic Skills (ITBS), typically takes 35 to 45 minutes per sub-test, depending on the grade level and the administration protocol. Typically, about 10 minutes of that time is taken up with proctor's instructions. Students with disabilities sometimes require extra time to complete the test. The full battery of the ITBS normally consists of four such subtests. It may be possible to administer two subtests for the evaluation, although subtests are often given separately, over consecutive days.

School districts that wish to be considered for the federal demonstration program may be willing to either provide adequate test data or make their students available for achievement testing. Nevertheless, under the study options with large numbers of schools, obtaining cooperation from each of the schools in the study could prove time-consuming and expensive and might still result in some schools not participating.

One alternative, testing in the school after school hours, is unattractive because students might be fatigued from a full day of school. Another alternative, testing on the weekends, would present its own problems. Probably the only feasible place to hold a weekend testing session would be the school itself, since obtaining space with furniture appropriately sized for different elementary school age groups would be difficult anywhere else. Therefore, proctors and building maintenance staff would have to be hired for the weekend session, adding costs to the evaluation. More important, inducing parents to bring their children to a weekend achievement test would be very difficult and would risk a high nonresponse rate, even if the evaluation were to provide significant incentives to parents and children. An advantage to weekend testing is that one can have a single session for students in multiple schools. This would save costs in the design options that call for smaller numbers of students spread out over many schools in the same SFA. Rather than pulling out a handful of students in each school, the evaluation could invite a large group of students from around the district to come to a single site.

In most of the discussion here, it has been assumed that tests would be administered to a group of students rather than to individuals. Indeed, most standardized achievement test protocols are written for group administration. There are some tests, however, that may be feasible and even desirable to administer individually. These would include computer adaptive tests, such as those recently developed for elementary students by CTB/McGraw Hill Publishing. These tests are currently being used in Minneapolis. Mentioned briefly in Chapter IV, computer adaptive tests would have each student work on a computer one at a time. If several computers were available, then testing a group could take a similar amount of time as that required for normal paper-and-pencil tests. Since computer adaptive tests have the potential for measuring achievement accurately in a much shorter time period than test booklets, this method of administration might actually save time. Students could take the tests in batches and they could make up a missed test at any time. Proctors

would need little training. In addition, the innovation of introducing adaptive tests might appeal to teachers and encourage cooperation with the study.

Another important consideration for collecting student achievement data is the lack of incentives for students to perform on the tests. It may be necessary to work with the school district to find solutions to this problem, including the reporting of individual scores to families. Anecdotal evidence from Minneapolis suggests that students enjoy the computer adaptive tests enough to make an effort to do well. Adaptive tests typically generate less frustration with questions that are too difficult or too easy because the items each student faces are chosen to provide the optimal amount of challenge. Therefore, students taking an adaptive test should be less likely to give up and more likely to take the test seriously.

Finally, one might need to consider test data collection logistics when deciding whether to test all children or a subsample. A subsample could be a sample of children in one or two grades, or a sample of children across all grades. Issues of vertically equated scales and potential aggregation bias are discussed in Chapter IV. Here, it is important to note that there are potentially high fixed costs per grade. That is because one normally needs separate booklets, proctors, scoring, and even separate rooms for each grade level. Separate rooms per testing session are necessary because proctors may be reading different instructions to the group and because early elementary grade students often require smaller chairs and desks than older elementary students. Taken together, the conceptual and practical reasons would suggest that selecting one or two grades is more efficient.

If the evaluation were to use a subset of two or more grade levels, there are potential cost savings in the particular grades chosen. It may be most cost-effective to include consecutive cohorts rather than spaced cohorts. That is because the test form used for a pre-test for one grade could serve as a post-test for another grade. For example, measuring achievement growth over grades 3 and 4

would require only three test forms, those for grades 2, 3, and 4. Measuring growth over grades 3 and 5, would require four forms, for grades 2, 3, 4, and 5.

There are two important qualifications to this argument, however. First, if computer adaptive tests were used, many of the fixed per-grade costs, as well as the aggregation problems would be less of a concern. For instance, children of all grades, ages, or ability levels could be tested together, since the individual items would be customized, and the test could draw items from a very large item bank covering a wide range of achievement (and hence age and grade) levels. Second, by sub-setting grades the demonstration would lose the ability to generalize to the other grades. Choice of which grades to include would be somewhat arbitrary from a policy perspective. It would be hoped that the intervention could raise achievement levels for students at all levels.

b. Timing of Data Collection

As Chapter IV discusses, if district-administered achievement tests are available and adequate, the data collection plan calls for collecting baseline (spring 2000) scores, as well as collecting subsequent district-administered test scores on an ongoing basis throughout the evaluation. If administered as part of the evaluation, however, a follow-up academic achievement test would be conducted in spring 2001 on the same schedule, although not on the same day, as the collection of other student outcome data from cognitive testing, student surveys, and the dietary intake interview.

4. Same-Day Cognitive Tests

Implementation of the USBP (and breakfast program participation, generally) may influence not only students' long-term academic achievement but also their short-term memory and cognition and behavior. Developmental psychologists have produced a wide array of assessment instruments

covering many domains of possible interest to the study, including interviews and checklists designed for parents and teachers, as well as tasks and tests administered directly to children.

a. Methods

The USBP evaluation will include administration of a short-term test of cognition--the Wechsler Memory Scale, and student emotional behavior--the Revised Children's Manifest Anxiety Scale (see Chapter IV). Typically, these tests take much less time to administer than standardized achievement tests (see Table IV.2). They will be administered to the child by the evaluation team at school.

b. Timing of Data Collection

The cognitive tests will be conducted on the same day that the students' dietary intakes are measured, after breakfast and before lunch is consumed. It is preferable to conduct the cognitive test before administering the 24-hour dietary recall. Table VI.2 shows that this testing should be administered in spring 2001.

c. Issues

Two issues of timing are important with respect to the administration of cognitive tests. First, the tests should be administered on the same day that students' breakfast program participation status and dietary intake at breakfast are measured, perhaps in the late morning. Second, since students' state of mind should be as normal as possible at the time they take these tests, the tests should be administered before the unusual data collection activities take place. For the same reason, it would be best if the cognitive tests were given to students in their schools, so as to minimize the disruption caused by the test-taking.

D. DATA COLLECTION FROM TEACHERS (ON THE SAMPLED STUDENTS)

The value of conducting a survey with the teachers of sampled students is that it can yield additional information about students in the school and the general school climate. The teacher data collection will have three components. Part one of the questionnaire will be designed to gather information about teachers' characteristics, their impressions of the school climate and the students in the school, and their perceptions of how implementation of the USBP has affected their schools.[4] This component would also obtain information about the classes in which the sampled students are enrolled, including classroom characteristics (ability level, climate, and so on) for the class attended by each student sampled. Teachers can also provide information about the length of the class, the number of times it meets, and the proportion of class time actually dedicated to instruction and learning.

The second part of the teacher questionnaire will be designed to include questions capturing some school outcome information for individual students in the teacher's class. Each teacher would be expected to complete one of these modules for each sample student in his or her class, reporting on students' usual attendance, tardiness, attentiveness, classroom behavior, health, academic strengths and weaknesses, and other factors affecting classroom performance.

A third component of the teacher data collection would involve ratings of students' usual cognitive and emotional functioning. A standardized teacher rating scale for assessing classroom behavior and attention--the Connors Teacher Rating Scale--will be administered to teachers. For example, the 10-item Hyperactivity Index of the Connors Teacher Rating Scale (CTRS-39) has been associated with school breakfast participation level, school breakfast participation change, and child

---

[4]For teachers in control schools, this set of questions would be replaced by another set that asks about the regular SBP.

hunger (Murphy et al. 1998a; and Murphy et al. 1998b). Further discussion of the use of the CTRS appears in Chapter IV.

1. Methods

In order to have teachers report for individual students, a crosswalk between sampled students and teachers should be planned and incorporated into the initial development of the student sampling frame. Although most public elementary schools are organized primarily into self-contained classes for instruction, some are departmentalized. Therefore, one manageable way to implement this type of teacher survey would be to specify which teacher (for example, English, math, first period) would respond to a short series of questions about classroom behavior for each sample member. The burden on individual teachers in departmentalized schools could potentially be greater. This technique has worked, however, in other studies (for example, in the U.S. Department of Education's National Education Longitudinal Study of 1988, Teacher Questionnaire-NELS: 88 Base Year).

The teacher questionnaire regarding school and classroom climate, general behavior of the sampled students, and the CTRS would be distributed to teachers, with followup conducted by phone.

2. Timing of Data Collection

As Table VI.2 shows, data collection for the teacher survey would be conducted during spring 2001. More specifically, the teacher rating scale should be conducted in conjunction with the school breakfast program participation and dietary intake--also in spring 2001.

3. Issues

The study must plan in advance which teachers will be completing both the teacher questionnaire and the teacher rating scale. Teachers may be reluctant to report individual-level student behavior without parental consent.

E. OBSERVATION OF SCHOOL MEALS

Because usual school breakfast participation is such an important variable to the evaluation, direct observation of school breakfast for sample students will be conducted for a target week. This would provide a direct measure of usual USBP/SBP participation in the school for sample students; and would supplement data on a target day and usual participation provided by the student/parent survey.

1. Method

Field interviewers, or perhaps school staff, would observe students as they passed through the cafeteria line or ate in the classroom, and check the names of students who select any foods from the school breakfast. Alphabetical student lists by classroom or for the study sample would be used to record the students who selected foods from the school breakfast.

b. Timing of Data Collection

Observation of school breakfasts would be collected for the target week (five days) that coincided with the 24-hour dietary data collection. Using this approach, the observed USBP/SBP participation would be used to verify the student's reported participation on the target day. The information would also be used to calculate the weekly USBP/SBP participation rate and to compare

this information to the parents' report of usual participation over the school year. School meals for sample students would be observed in spring 2001 and spring 2002.

c. Issues

Observation of school meals would require another type of data collection in the school, but it could be accomplished at the same time as the completion of the Menu Checklist and the a la carte checklist (see Chapter VIII). Observation of students' school breakfast would also require parental consent.

VII. IMPACT ANALYSIS

A central aspect of the USBP demonstration will be to analyze data that have been collected to address the key questions of the evaluation. This chapter of the report defines the analysis questions to be addressed and outlines the basic estimation procedures through which these questions can be addressed. Understanding the analytic strategy to be used sheds light on the implications of making particular decisions about sampling and measurement issues.

The rest of this chapter contains four sections. Section A introduces the types of impact analysis to be conducted and describes the objectives of each type. Section B discusses the trade-off between using individuals and using schools as the unit of analysis. Section C presents the models to be estimated, and Section D touches on the key issue of selection bias.

A. TYPES OF ANALYSIS

According to the Child Nutrition Act of 1998, a key aim of the analysis in the USBP demonstration evaluation is to "evaluate the effect of providing free breakfasts to elementary school children, without regard to family income, on participation, academic achievement, attendance and tardiness, and dietary intake over the course of a day." Three major types of analysis can help achieve this objective:

1. Descriptive analysis
2. Estimation of the effects of USBP availability on the student outcomes listed above
3. Estimation of the effects of SBP/USBP participation on these student outcomes

1. Descriptive Analysis

The objectives of the descriptive analysis are to provide a context to the estimation of program effects and to address specific methodological issues related to the evaluation design. The descriptive analysis involves estimating and comparing the mean characteristics or outcome measures of various groups of students. Three sets of comparisons are especially relevant:

1. Students at USBP schools versus those at regular schools
2. Breakfast program participants versus nonparticipants
3. USBP participants versus regular SBP participants

Comparing the characteristics of students in USBP schools with those of students in regular SBP schools will serve as a check on the effectiveness of random assignment. If random assignment works as designed, the characteristics of the two groups will be similar. Analogously, comparing the characteristics of breakfast program participants versus nonparticipants will reveal the extent to which the decision to participate is related to observable characteristics. Since students are not randomly assigned into participant and nonparticipant groups, it is not expected that the characteristics of the two groups will be identical. However, large differences in the observable characteristics of the two groups would suggest that the groups may also differ in unobserved ways, implying that selection bias is a problem in estimating the effects of participation on student outcomes.

The degree to which students' characteristics are related to their participation decision may differ for students in USBP and regular SBP schools. One way of checking to determine whether this is true involves comparing the characteristics of USBP participants with those of regular SBP participants. If the characteristics of these two groups differ greatly, one could conclude that the

extent to which there is selection by students into the USBP differs from the extent to which there is selection into the regular SBP. Furthermore, this would imply that the two different breakfast programs serve different groups of students. This part of the descriptive analysis would better help understanding of the differences in the populations being served.

A final type of descriptive analysis not mentioned above would involve describing the characteristics of the districts participating in the evaluation. Although the evaluation is designed to be internally valid only, comparing the characteristics of the participating school districts with those of other districts nationally will lead to better understanding of the degree to which the evaluation results are *suggestive* of what the effects of the USBP would be if it were implemented nationally.

2 Estimation of the Effects of USBP Availability on Student Outcomes

A key component of the USBP demonstration evaluation will be to estimate how the implementation of the USBP in a given school influences outcomes among students at that school. As mentioned above, random assignment of schools into USBP and regular SBP statuses implies that a simple comparison of mean outcomes among students at USBP and regular SBP schools serves as an estimate of the effect of USBP availability on these outcomes. However, estimating this effect after controlling for other student characteristics in a regression framework yields a more precise estimate of this effect, with no loss in experimental rigor.

Since the aim of this analysis is to estimate the effect of USBP *availability* on student outcomes, it is important to realize that this estimated effect would be based on a sample of all students attending USBP and regular SBP schools, including those who would participate in either the USBP or regular SBP and those who would participate in neither program. Since USBP availability is unlikely to have a large effect on student outcomes among those who do not change their

participation status in the breakfast program, the large number of these students who do not change their participation behavior after the USBP is implemented is likely to dilute the estimated effects of program availability, making them more difficult to detect statistically. In other words, USBP availability may have a strong effect among the small number of students who consider participating in the breakfast program, but its overall effect may be small (and statistically insignificant) because most students either do not consider participating in the program or would participate in the program with or without the USBP.

The strengths of the estimation of the effects of USBP availability on student outcomes are that it takes advantage of the experimental design and yields estimates that are policy-relevant. In particular, it addresses the following policy question: "If the USBP were to be implemented in a given school, what changes would result in the likelihood of skipping breakfast, breakfast program participation, dietary intake, school attendance, academic achievement, and other outcomes among all students at that school?" Because the evaluation is not designed to be externally valid, the answer to this question will be valid only for schools in the demonstration districts. However, the answer will be suggestive of how the USBP would affect student outcomes for all districts.

This analysis should be conducted on the full sample and on key subgroups of the full sample. One set of subgroups of particular interest is defined by income level. Poor students (defined as those whose income is below either 185 percent or 130 percent of poverty) may be affected differently than nonpoor students by a USBP for two major reasons. First, assuming that a large proportion of poor students are certified for free meals, the shift from the regular SBP to the USBP would not affect the price these students pay for breakfast; they would pay nothing in either case. However, poor students may be more likely to participate in the USBP than in the regular SBP because of the elimination of stigma and because a larger number of other students may be

participating. Second, the families of poor students may be more likely than the families of wealthier students to view free breakfasts at school as a substantial benefit. While the families of most nonpoor children probably do not have to struggle much to feed their children, the families of some poor children are likely to be spending a larger proportion of their budget on food, so a free meal at school may help them obtain other necessities. It is not clear whether poor children or nonpoor children would be more dramatically affected by the USBP, but it will be important to test for differences between the two groups.

3. Estimation of the Effects of Breakfast Program Participation on Student Outcomes

The objective of the third type of analysis is to determine how breakfast program *participation*--that is, how eating a school breakfast--influences student outcomes such as dietary intake and academic achievement. This is an important objective, for three reasons. First, a basic premise of the USBP demonstration is that promoting breakfast program participation will help improve student outcomes. Estimating the actual effects of participation will serve as a check on this premise of the value of participation. Second, USBP availability presumably influences student outcomes through a two-step process: it promotes breakfast program participation, which, in turn, improves various student outcomes. If this model is correct, the estimates of the effect of participation on student outcomes will make possible the decomposition of the overall effect of USBP availability into these two components. Third, an important effect of implementing the USBP may be to change the way breakfast program participation influences student outcomes. Since the analysis of the effects of participation on outcomes can be conducted separately among students in USBP and regular SBP schools, it will indicate whether the two types of programs have different effects.

The analysis should include estimation of both the short-term and long-term effects of participation. The key short-term effects of interest include whether participation on a given day

influences students' dietary intake and cognitive functioning on that day. The key long-term effects of interest include whether students' usual participation influences their academic achievement and school attendance/tardiness rates.

The experimental design of the USBP evaluation will not guarantee unbiased direct estimates of the effects of participation on student outcomes. Without random assignment of individual students into participant and nonparticipant groups, the methodology for estimating the effects of participation on student outcomes will involve estimating a set of regression equations in which the dependent variables are the student outcomes of interest, and the independent variables include the student's breakfast program participation status and relevant student and school characteristics. Even after controlling for observable characteristics, participants and nonparticipants may differ in unobservable ways. If these unobserved differences are not taken into account, the estimates of the effects of participation may be biased. Thus, the models for estimating these effects may have to incorporate some type of correction for these unobserved differences (that is, selection bias). This methodological problem makes the analysis design for the estimation of the effects of participation on student outcomes weaker than the analysis design for the estimation of the effects of USBP availability on student outcomes. Section D describes this issue in more detail.

B. UNIT OF ANALYSIS

The analyses described above could be conducted at the school level or the student level. In other words, the dependent variables being analyzed (along with the key independent variables being analyzed) could represent outcomes for specific individuals (student-level analysis) or represent average outcomes among students at specific schools (school-level analysis). Each approach has advantages and disadvantages (as described below). Therefore, the analysis should address the key

research questions at *both* the school level and the student level whenever possible. This will provide additional evidence regarding the questions central to the evaluation.

1. School-Level Analysis

If schools are used as the unit of analysis, both the student outcome measures and the key characteristics will be measured at the school level. For example, key outcomes might include the breakfast program participation rate at the school, mean test scores among students at the school, and the school's attendance and tardiness rates. Key independent variables in the school-level analysis will include the school's USBP status, school characteristics such as size and location, and average characteristics of students at the school.

The main drawbacks of conducting the analysis at the school level involve the precision of the resulting estimates. Since the sample of schools is relatively small, the standard errors of the estimated effect of USBP availability (and other independent variables in the model) may be relatively too large to reliably estimate most outcomes other than breakfast program participation (see Section V.C). In addition, variation in the outcome measures and some of the key independent variables will be lost if these variables are measured at the school level rather than the student level. For example, information on which students participate in the breakfast program and which do not will be lost in a model whose dependent variable is the mean participation rate in the school. On the other hand, for characteristics (such as the school's USBP status) that vary only from school to school and not among students at a given school, no such variation will be lost. For this reason, the school-level analysis is more appropriate for estimating the effect of USBP availability on student outcomes than it is for estimating the effect of breakfast program participation on student outcomes.

The main advantage of the school-level analysis is that it is usually easier to collect data at the school level than at the student level. For example, schools may freely report their SBP participation

rate, mean test scores, and average attendance rate. By contrast, data on individual students' participation status, test results, and school attendance may be more difficult to obtain. Moreover, school-level data may be available for multiple years, allowing the analysis to take advantage of variation in the dependent and independent variables over time, as well as across schools.

2. Student-Level Analysis

In the student-level analysis, outcome variables are measured at the student level, and most of the independent variables are also measured at the student level. Other independent variables (such as a school's USBP status) are defined at the school level but then are applied separately to each individual in that school.[1] The student-level analysis can be used as easily as the school-level one to estimate the effects of USBP availability and breakfast program participation on student outcomes.

The major advantage of the student-level analysis over the school-level analysis is that the sample size is much larger, so there is more variation in the outcome measures and independent variables. Thus, if information is available on comparable numbers of students, the precision of the student-level estimates typically is greater than the precision of school-level estimates. The major drawback of the student-level analysis is that student-level data are sometimes costly to collect.

C. MODELS TO BE ESTIMATED

As described in Sections VII.A and VII.B, the analysis will include models that estimate the effects of USBP availability and breakfast program participation on student outcomes. The models

---

[1]This multilevel measurement of characteristics leads to multilevel, or hierarchical linear, modeling (as described in Section VII.C).

estimated will use either student- or school-level data. This section presents the econometric models that can be used in each of these cases.

1. Model I: School-Level Model of the Effects of USBP Availability

The following model can be used to estimate the effects of USBP availability on student outcomes, using school-level data:

$$Y_j = Z_j\delta_1 + \alpha_1 USBP_j + \epsilon_{1j},$$

where

Y_j = mean outcome among students at school j
Z_j = vector of characteristics of school j
$USBP_j$ = binary variable representing USBP status of school j.

In this model, an outcome measured at the school level is hypothesized as depending on school characteristics and the school's USBP availability. The dependent variable in the model may be the school's breakfast program participation rate, attendance rate, mean achievement test scores, or some other school-level measure. The control variables in the Z_j model include relevant and observable school characteristics such as enrollment and racial composition. For a full list of potential dependent and independent variables, see Chapter IV. Depending on whether the dependent variable is continuous, binary, or truncated, Model I could be estimated using OLS, logit/probit, or tobit estimation techniques.[2]

---

[2]Since the sample is intended to be representative of the districts in the demonstration, and it is unlikely that the evaluation will sample every elementary school in these districts, the sample will be clustered. Thus, individual-level observations among students within the same school are not likely to be statistically independent due to these design effects. Furthermore, it is likely that pairs of similar schools will be selected before one is randomly assigned to the USBP group and the other to the regular SBP group. This random assignment process will lead to correlation in the characteristics of treatment and control schools (as well as the students in these schools), further

(continued...)

Estimation of this model will produce estimates of the parameters δ_1 and α_1, with the latter being the estimate of the effect of USBP availability on the outcome of interest. Random assignment of schools into the USBP and regular SBP groups ensures that the estimate of this effect is unbiased.

This simple school-level model can easily be modified in at least two ways in different situations. First, if the effects of USBP availability were thought to differ in different types of schools, then $USBP_j$ could be interacted with the relevant school characteristics in Z_j. Second, to increase the precision of the model estimation, data from more than one year could be used, if available.

2. Model II: Student-Level Model of the Effects of USBP Availability

The following model can be used to estimate the effects of USBP availability on student outcomes, using student-level data:

$$Y_{ij} = X_{ij}\beta_2 + Z_j\delta_{2j} + \alpha_{2j}USBP_j + \epsilon_{2ij},$$

where

$$\delta_{2j} = d_2 + u_{2j}$$

$$\alpha_{2j} = a_2 + e_{2j},$$

where

Y_{ij} = value of outcome among student i at school j
X_{ij} = vector of characteristics of student i
Z_j = vector of characteristics of school j
$USBP_j$ = binary variable representing USBP status of school j.

---

[2](...continued)
leading to correlation across individual- and/or school-level observations. Thus, the estimated standard errors of the coefficients will have to be adjusted to account for the correlation across observations. Taylor series approximation methods, such as those implemented in the SUDAAN statistical software, can be used for making these adjustments, which typically increase the magnitude of the estimated standard errors.

This is a multilevel model, or hierarchical linear model (HLM), where the value of the outcome measure (for example, nutrient intake or achievement test score) is hypothesized as depending on both student and school characteristics.[3] However, the effects of school characteristics on the student outcome are allowed to vary across schools. In the specification shown here, the effects of the school characteristics included in Z_j include a constant factor (d_2), which is the same across all schools, and a random factor (u_{2j}), which varies across schools. Similarly, the effect of USBP status includes a constant (a_2) and random factor (e_{2j}). At this second (school) level, one could also have specified that the effects of Z_j and $USBP_j$ vary systematically across schools, depending on the characteristics of those schools. For simplicity, the term is not included. It is also assumed that u_{2j} equals 0 for each of the components of δ_{2j} except for the component associated with the constant term in Z_j (we label this random school effect term v_j). With these simplifying assumptions, substituting the two school-level equations into the student-level outcome equation yields:

$$Y_{ij} = X_{ij}\beta_2 + Z_j d_2 + a_2 USBP_j + v_{2j} + e_{2j} * USBP_j + \epsilon_{2ij}.$$

In this equation, the last three terms represent the error structure of the model.[4] This multilevel specification allows us to account for two potential, problematic aspects of the data. First, the model explicitly models the potential nonindependence of the error terms across observations, which is likely because the sample includes multiple students from the same schools. Second, the

---

[3]This model could also be used to estimate the effects of an intervention in which certain students within a school are randomly assigned to receive a free breakfast and others are randomly assigned to receive a reduced-price breakfast. In this case, the variable $USBP_j$ would be replaced by a variable ($FSBP_{ij}$) that would indicate whether or not a given student was offered a free breakfast.

[4]An even simpler version of this model would eliminate the term $e_{2j} * USBP_j$, which would eliminate the heteroskedasticity from the model. This could be justified either by eliminating the random element from the coefficient α_{2j} or by assuming that the variance of v_{2j} is very large relative to the variance of e_{2j}.

specification allows for a systematic relationship between the values of the independent variables ($USBP_j$, in this case) and the variance of the overall error term. Straightforward techniques exist for the estimation of this multilevel model (see Bryk and Raudenbush 1992).

The key estimate from the model is the estimate of the parameter a_2, which represents the average effect of the availability of the USBP in a given school on the student outcome of interest. Again, random assignment of schools into USBP and regular SBP statuses ensures that this estimate will be unbiased. If the outcome measure is students' participation in the breakfast program, a_2 will represent the direct effect of USBP availability. For other outcome measures such as dietary intake or test scores, however, a_2 will represent primarily an indirect effect. The main way in which USBP availability is expected to influence outcomes like these is by first influencing program participation, which, in turn, influences dietary intake or academic achievement.[5]

A couple of unique aspects of this model arise when the dependent variable is students' performance on a standardized test. First, the vector of student characteristics (X_{ij}) will likely include a lagged value of the test score dependent variable.[6] With the inclusion of this variable, the dependent variable effectively measures the gain in students' test performance from one year to the next, rather than the absolute level of their test performance. Compared to the test score level

---

[5]The possibility exists, however, that USBP availability has a direct influence on such outcomes as intake or test scores. For example, implementation of the USBP may affect the quality of breakfasts served for all students, which has a direct effect on intake even if the overall participation rate does not change.

[6]The inclusion of a lagged value of the dependent variable as an independent variable will result in errors in variables problem (Greene 1993). This is particularly likely to be true when the dependent variable is test performance. However, with a known reliability coefficient associated with the test (which is often published), making a measurement error correction should be straightforward.

variable, the test score gain variable should have lower variability and be more sensitive to policy changes in place only during the past year.

Second, it is likely that different standardized tests will be administered in different districts included in the demonstration. Under certain circumstances, however, it will be possible to aggregate observations of students taking different standardized tests. In particular, the two requirements for aggregating across different tests are (1) that the tests measure the same underlying construct (for example, math or reading skills); and (2) that the tests are normed to similar populations, such as schoolchildren nationally at about the same time period, and hence scaled to the same units.

3. Model III: School-Level Model of the Effects of Breakfast Program Participation

In all likelihood, estimating a school-level model of the effect of breakfast program participation on student outcomes will be useful only if the data include multiple years of information on school-level participation rates, school characteristics, and outcome measures. With only a single year of data, sample sizes are too small, and too much information is lost on the participation status of individual students' within the same schools. With multiple years of data, the school-level model is:

$$Y_{jt} = Z_{jt}\delta_3 + \alpha_3 P_{jt} + \gamma_{3j} + \epsilon_{3jt} .$$

where:

Y_{jt} = mean outcome among students at school j in year t
Z_{jt} = vector of characteristics of school j in year t
P_{jt} = participation rate among students at school j in year t
γ_j = fixed effect of school j.

The key parameter in the model is α_3, which represents the influence of the participation rate in the school on outcomes like the school's mean test score or overall attendance rate. The term γ_j represents the fixed effect of a given school, or the overall influence on the outcome of unobservable factors associated with that particular school. Without random assignment, the school participation rate may be correlated with this fixed effect or with the time-varying error term (ϵ_{3jt}).

A key, implicit assumption of the model specified above is that the effect of participation on outcomes such as these is the same in USBP schools as in regular SBP schools. This assumption is maintained here for simplicity, but it could easily be relaxed by interacting the participation rate with the binary variable indicating whether the USBP is available in a school (in fact, the assumption is relaxed in Model IV).

For this model to be estimated with a reasonable degree of precision, the school-level participation rate must have sufficient variation over time and across schools. This is likely to be the case in a model in which the effect of participation is assumed to be the same in USBP and regular SBP schools, since the participation rate is likely to be much higher in USBP schools. However, if this assumption is not made, and USBP and regular SBP participation are treated differently, it is not clear that these two participation rates will vary greatly over time or across schools. This must be carefully monitored.

An advantage of the model as specified above is that it is possible to directly control for the unobserved school fixed effect (γ_j) using binary variables representing each school in the sample. As long as there are multiple years of data from the same schools, it is possible to control for the unobserved time-invariant characteristics of each school that influence the outcome. The result is a model whose key estimated parameter indicates the extent to which a school's attendance rate or test scores change when there is a change in the participation rate in the school (controlling for how

these outcomes change in schools with no change in the participation rate). Thus, to estimate the model with a reasonable degree of precision, there must be exogenous changes in the participation rates of individual schools over time.

4. Model IV: Student-Level Model of the Effects of Breakfast Program Participation

The following model can be used to estimate the effects of breakfast program participation on student outcomes, using student-level data:

$$Y_{ij} = X_{ij}\beta_4 + Z_j\delta_{4j} + \alpha_{4j}P_{ij} + \epsilon_{4ij},$$

where

$$\delta_{4j} = d_4 + u_{4j}$$
$$\alpha_{4j} = a_4 + b_4 USBP_j + e_{4j},$$

where

Y_{ij} = value of outcome for student i at school j
X_{ij} = vector of characteristics of student i
Z_j = vector of characteristics of school j
P_{ij} = variable indicating the participation status of student i.

In this model, the definition of P_{ij} differs, depending on whether the dependent variable is a short-term or long-term outcome. If it is a short-term outcome, like dietary intake or cognitive functioning, P_{ij} is a binary variable that measures whether or not the student eats a school breakfast on a given day. If the dependent variable is a long-term outcome like the student's test score gain or attendance rate over a school year, P_{ij} represents the student's usual participation. For example, it may be defined as the number of days per week that the student usually eats a school breakfast.

This HLM model is similar to Model II, except that the second school-level equation is slightly more general. This equation allows the effect of participation on the outcome measure to vary by

school, not only randomly (through e_{4j}) but also systematically through the school's USBP status. In other words, the model allows the effect of eating a school breakfast to differ for students in USBP schools, versus students in regular SBP schools. Again, assuming that all the components of u_{4j} equal zero except for the component associated with the constant term, the model reduces to the following:

$$Y_{ij} = X_{ij}\beta_4 + Z_j d_4 + a_4 P_{ij} + b_4 P_{ij} * USBP_j + v_{4j} + e_{4j} * P_{ij} + \epsilon_{4ij}$$

In the model, the average effect of participation on the outcome among students in regular SBP schools is represented by the parameter a_4, while the average effect of participation among students in USBP schools is represented by the parameters $(a_4 + b_4)$. If the coefficient estimate of the parameter b_4 is statistically significant, this indicates that the effect of participation in USBP schools differs significantly from that in regular SBP schools.[7]

For the usual HLM estimation techniques to lead to unbiased estimates of a_4 and b_4, the model must include the assumption that P_{ij} is not correlated with the error terms in the model. In other words, the unobserved school and student characteristics affecting the outcome must not be related to whether or not students eat a school breakfast. Since individual students are not randomly assigned to participant and nonparticipant categories, the experimental design of the overall study *does not* ensure that this assumption is true. In fact, a variety of factors lead us to question this

---

[7]It is also possible that the effect of participation is nonlinear--that is, the effect changes as more and more students participate--or that there are additional interactions to the participation effect besides the interaction with the school's USBP status. For example, the effect may be larger for low-income students than for high-income students. These possibilities can be explored through specification tests of Model IV.

assumption. If the assumption is not true, selection bias in the estimation of a_s and b_s will result. This issue is discussed in the next section.

D. SELECTION BIAS

As long as schools are randomly assigned to USBP and regular SBP status, there will be no self-selection of schools into USBP status and no resulting selection bias in the ordinary least squares (OLS) estimates of the effect of USBP *availability* on continuous student outcomes. Within schools, however, there will be self-selection of students into breakfast program participant versus nonparticipant status. In other words, unobserved student characteristics will influence whether or not they become breakfast program participants. For example, students' home situation, unobserved aspects of their socioeconomic status, or their dietary habits and attitudes may influence their participation decision. If these factors also influence the key outcome variables, then estimates of the effect of participation on the outcomes that do not take these unobserved factors into account will suffer from selection bias.

The ideal solution to this selection bias problem is to try to ensure that it does not arise at all, by controlling for all relevant factors that might influence the outcome measure. If this does not work, two possible methods for accounting for the unobserved differences between participants and nonparticipants are (1) to estimate instrumental variables (IV) or selection correction models, and (2) to estimate fixed-effect models. These approaches are not mutually exclusive; some combination of them can be used to address the issue of selection bias. A third approach is to attempt to control for the unobserved factors that influence short-term student outcomes and are related to their participation status by including both participation status on a given day and usual participation in the same model.

Each of the approaches for dealing with selection bias in the estimation of participation on student outcomes has weaknesses. Thus, estimates of the effect of participation on these outcomes from any one of these models may be biased. The best overall estimation strategy involves using a variety of different approaches to control for selection bias and to compare the resulting estimates of the effect of participation. If the different approaches lead to similar results, they will be more believable.

1. The Problem

The problem of selection bias can be illustrated econometrically using a two-equation model. The first equation is similar to the student-level model showing the effects of participation presented in the previous section, except that a time subscript is added to represent the model at two different points in time. In addition, three simplifying assumptions are made: (1) that the effect of participation is the same in USBP and regular SBP schools, (2) that there is no interaction between the error term and any of the independent variables in the model, and (3) that there is no random school effect. In addition, an individual fixed effect to represent unobserved individual characteristics that influence the outcome but do not vary over time can be added to the specification. The second equation is a participation equation:

$$Y_{ijt} = X_{ijt}\beta_4 + Z_{jt}d_4 + a_4 P_{ijt} + (\gamma_{4i} + \epsilon_{4ijt})$$

$$P_{ijt} = W_{ijt}\phi_4 + V_{jt}\psi_4 + (\theta_{4i} + \mu_{4ijt}).$$

In the model, an outcome measure (Y_{ijt}) is regressed on student characteristics (X_{ijt}), school characteristics (Z_{jt}), and participation status (P_{ijt}), with the error structure including an individual-

specific fixed effect (γ_{4i}) and a random error term (ϵ_{4it}). The second equation of the model shows participation status being regressed on student characteristics (W_{it}) and school characteristics (V_{it}), with the error structure also including an individual-specific fixed effect (θ_{4i}) and random error term (μ_{4it}).[8]

In the model, selection bias arises if participation status is correlated with either error term in the first equation. In particular, if either term in the participation equation error structure ($\theta_{4i}+\mu_{4it}$) is correlated with either term in the outcome equation error structure ($\gamma_{4i}+\epsilon_{4it}$), selection bias will result. Another way of saying this is that selection bias will arise if any of the unobserved factors (either time-invariant individual-specific factors or time-varying random factors) that influence whether a person eats a school breakfast are related to any of the unobserved factors (again, either fixed or time-varying) that influence the outcome of interest.

Below, four possible approaches for addressing the selection bias issue are discussed. In evaluating the proposed approaches, it is important to distinguish between short-term and long-term models. Short-term models involve outcomes that are measured on a single day, such as dietary intake and cognitive functioning, and are presumed to be influenced by whether the student eats a school breakfast on that day. Long-term models involve longer-term outcomes, such as performance on standardized tests or student attendance over a full school year, that are presumed to be influenced more strongly by students' usual SBP participation than by their participation status on a single day. The unobserved factors affecting the short-term outcomes are likely to differ from those affecting long-term outcomes. Similarly, the unobserved factors related to participation on a single day are likely to be different from the unobserved factors related to usual participation.

---

[8]If the model is being estimated based on only a single year of data, the "t" subscript can be dropped.

2. Controlling Directly for Relevant Factors

In the model described above, if all of the relevant student and school characteristics that influence the outcome variable are measured and included in X_{ijt} and Z_{jt}, then the error term ($\gamma_{4i}+\epsilon_{4ijt}$) will be truly random and not correlated with the error term in the participation equation ($\theta_{4i}+\mu_{4ijt}$).[9] In this case, there will be no selection bias and the simple OLS estimate of a_4 will be an unbiased estimate of the effect of participation on the outcome of interest (assuming that Y_{ijt} is continuous and the model is specified correctly). If feasible, this is the best solution to the problem of selection bias because, compared with the other potential solutions described below, it requires fewer assumptions, is typically less sensitive to the exact specification used, and results in a more precise estimate of the effect of participation on the outcome of interest.

The primary challenge in implementing this solution is that many of the relevant factors influencing outcomes like dietary intake or academic achievement are difficult to measure. For example, factors such as students' dietary habits, their parents' dietary knowledge and attitudes, details about their socioeconomic status, and the characteristics of the food service operation at their school are not available in most data sets and are not easy to measure, but these factors potentially influence students' dietary intake and academic achievement.

However, given that the USBP evaluation will involve primary data collection, the possibility exists for designing data collection instruments to attempt to obtain this sort of information. Section G of Chapter IV describes the basic control variables to be collected and used in the models to be estimated. In addition to these variables, however, a few examples of other possibly relevant pieces of information that could be collected include:

---

[9]In this discussion, "relevant" characteristics or factors refer to those that have an effect on the outcome of interest and are also correlated with participation status (or any of the other variables in X_{ijt} or Z_{jt}).

- Detailed information on children's socioeconomic status
  - Family wealth and assets (for example, bank holdings, vehicle ownership)
  - Whether family has specific possessions in the home (for example, a personal computer)
  - Employment histories of all adults in the household
  - Information about the socioeconomic status of the family's neighbors
- Dietary habits of the sample member
  - Favorite foods
  - Foods they will not eat
  - Typical intake of the child on a weekend day
- Dietary knowledge and attitudes of food preparer in sample member's household
  - Knowledge of USDA Food Guide Pyramid servings recommendations
  - Attitudes about the importance of particular nutritional guidelines
  - Usual food shopping habits

If an outcome model could be properly specified and all the relevant control variables were included in it, none of the more complex econometric methods for dealing with selection bias would be necessary. However, if the data collection effort does not yield sufficiently rich and detailed information for this purpose, the methods described below may be useful.

3. Fixed-Effects Models

Estimating fixed-effects models involves finding a way to directly control for the individual-specific fixed, or time-invariant, part of the error term. This requires data on the outcome and participation status (as well as selected individual and school characteristics) from more than a single point in time.

For certain short-term outcomes, such as dietary intake, the required data would likely be collected for at least some students on 2 days, 3 to 10 days apart. With this information, a fixed-effects model could essentially be estimated by including a dummy variable for each individual in the relevant sample. By controlling for each individual, the model effectively controls for the unobserved characteristics of these individuals that are constant over time. Consequently, the fixed-effects estimate of the effect of participation on the short-term outcome is based on whether individuals who participate on one day, but not the other, have higher (or lower) dietary intake/cognitive functioning on the days in which they participate than on the days in which they do not, after controlling for observable factors that vary over time.

This type of fixed-effects model has two major weaknesses. First, most students will likely participate either on both of the two days or neither of the two days.[10] Thus, the effect of participation may not be estimated with a great deal of precision, since it is based on those individuals who participate on one day but not on the other. Second, selection bias remains a potential problem with fixed-effects models. The model effectively controls for correlations between the unobserved time-invariant determinants of participation (θ_{4i}) and the outcome of interest (γ_{4i}) but does not control for potential correlated between the unobserved time-varying determinants of participation (μ_{4ijt}) and the outcome of interest (ϵ_{4ijt}). In other words, the unobserved factors that lead a student to participate on one day but not on another day may be related to the unobserved factors that influence the student's dietary intake and/or cognitive functioning on those days. For example,

---

[10]Since information on students' school breakfast participation on two separate days can be determined from Continuing Survey of Food Intakes by Individuals (CSFII) data, it is possible to get an estimate of the frequency with which students participate on one day but not the other. Among students attending schools that offer the SBP in the CSFII, there are 413 with information from two school days. Of these students, 12 percent participated in the SBP on both days, 76 participated on neither day, and only 12 percent (48 students) participated on one day but not on the other.

whether or not a student's parents leave for work early on a given day may influence both the student's intake and whether or not the student participates in the breakfast program. When the parents leave early, they have little time to prepare a breakfast for the student, who consequently eats less at home on that day. These are the days on which the student gets a school breakfast. When the parents leave for work late, they prepare a large breakfast for the student at home, and he or she skips the school breakfast. The student's intake on both days may be about the same, making it appear as though the program has no effect. With the breakfast program, however, the student would have eaten less on days on which the parents left early for work.

For the long-term fixed-effects models, the two points in time for which data are needed typically are a year apart. For example, students' usual SBP participation, their attendance rate and achievement test scores, and a variety of other student and school characteristics would be measured over two separate school years.[11] The estimate of the effect of participation in this model is based on whether individuals who usually participate more in one year than in another year have higher (or lower) attendance rates and greater (or smaller) gains in test scores in those years in which their usual participation is higher than it was in the years in which their usual participation is lower, after controlling for observable factors that vary over these two years.

The weaknesses of the long-term fixed-effects model mirror those of the short-term model. Just as most students either participate or do not participate on each of the two days of the short-term model, most students have similar levels of usual participation in each of the two years of the long-term model. In addition, selection bias based on time-varying unobserved factors is a potential

---

[11]Actually, since what is of interest with respect to test scores is the gain in students' scores from one year to the next, three years of test score data are required to estimate a fixed effects model. Data from the first two years are used to measure test score gains at one point in time, and data from the second and third years are used to measure test score gains at the subsequent point in time.

problem with the long-term model. Students whose situation changes from one year to the next (in some unobservable way) may become more likely to participation in the breakfast program but may also experience changes in their attendance/achievement that are unrelated to their usual participation.

4. Instrumental Variables/Selection Correction Models

Estimating IV or selection correction models uses a two-stage approach to account for selection bias. In the first stage, the participation equation (shown above) is estimated. In the second stage, the equation with the outcome variable of interest as the dependent variable (shown above) is estimated, with either the predicted breakfast program participation status of the student, based on the first-stage equation (in the IV model), or the student's actual participation status and a selection correction term also based on the first-stage equation (in the selection correction model) included as independent variables.

In either the case of the IV or the selection correction model, identification of the model depends in practical terms on finding appropriate identifying variables and including them in the first-stage participation equation but not in the second-stage outcome equation. Thus, these identifying variables must be correlated with students' participation status but have no direct effect on the outcome measure that is the dependent variable in the second-stage equation. In practice, finding appropriate identifying variables is difficult. Furthermore, models of this type have often been criticized for not being sufficiently robust--that is, different specifications of the same IV or selection correction model produce different results. The two keys to estimating IV or selection correction models are to find good identifying variables and to conduct a variety of specification tests to ensure that the model's results are robust.

Potential identifying variables to consider for the estimation of an Instrumental Variable or selection correction model of the effect of breakfast program participation on student outcomes include:[12]

- ***Timing Considerations.*** The amount of time a student has at home in the morning before leaving for school and the amount of time between his or her arrival at school and the start of classes may influence whether or not the student eats a school breakfast. Variables capturing these factors and related information, such as the length of the student's commute to school, the start of the school day, whether a parent or sibling is available to serve breakfast to the student at home, and the availability of before-school activities, may be used to construct appropriate instrumental variables.

- ***Breakfast Price and Students' Certification Status.*** In regular SBP schools, these variables, together, would indicate the price individual students pay for breakfast. If the model controls for family income, it could be argued that price influences participation but not dietary intake. In USBP schools, the variables would work differently. Even in USBP schools where breakfast is free, certification status may affect the price of lunch and thus usual lunch program participation, and usual lunch participation may be related to SBP participation.

- ***"Planted" Identifying Variables.*** Because the evaluation team is designing the evaluation, it may be able to "plant" identifying variables. In other words, it may be able to take some action that will lead to higher breakfast program participation among a randomly selected subset of sample members but does not directly influence the outcome measures. An example of this would be if the evaluation team contacted a random subset of sample members by phone or mail in order to encourage SBP participation. Alternatively, an intervention could take place at the classroom level; perhaps the teachers of a random subset of classes could become involved in promoting student participation. Students might be more likely to be influenced by a classroom intervention because their friends would be receiving the same intervention. If the effort to promote participation was successful and if the intervention did nothing to directly influence breakfast intake or other outcomes, then a variable indicating whether a given student received the intervention could be used as an identifying variable.

- ***Welfare Status.*** In regular SBP schools, welfare status could be linked to SBP participation, for two reasons. First, it makes becoming certified administratively easier.

---

[12]The identifying variables suggested here include only student-level variables. Although some school-level variables might be appropriate identifying variables in principle, they could not explain variation in the participation statuses of students within a given school. Because of this and because of the relatively small number of schools likely to be in the sample, school-level identifying variables are unlikely to be effective in practice.

Second, it may indicate a particular family attitude toward stigma. As with breakfast price, the link with SBP participation in USBP schools would have to be an indirect link, operating through participation in the school lunch program.

As with the fixed-effects models, the instrumental variables/selection correction models have weaknesses and are not likely to generate estimates of the effect of participation that all readers would agree are unbiased. Except for the evaluator-planted identifying variables, however, collecting the information needed to construct these identifying variables is likely to be relatively inexpensive. Thus, as noted above, the wisest strategy may be to estimate IV or selection correction models but also to conduct a wide variety of specification tests and checks of robustness. The goal of this strategy would be to accumulate evidence about the effects of breakfast program participation that are consistent across models and specifications.

5. Controlling for Usual Participation

In models designed to estimate the effects of participation on a given day on a short-term outcome, another possible way to address the selection bias problem is to control for usual participation in the outcome equation. In particular, an outcome such as dietary intake or cognitive functioning would be regressed on student and school characteristics, participation status on the day in which the outcome is measured, and usual participation. The estimate of interest is the effect of participation on that single day on the outcome. With one-day participation status already in the model, the variable measuring usual participation captures unobserved characteristics unique to individual students (and time invariant) that influence their participation generally (in other words, the time-invariant, individual-specific fixed effect).

This model is similar to the fixed-effects model in that it depends on variation in students' single-day participation after controlling for their usual participation. In other words, among

students who say they usually participate a given number of days a week, there must be some who participate and others who do not on the day on which the outcome variable is measured. In addition, this model leaves the possibility of selection bias based on unobserved time-varying characteristics. However, controlling for students' usual participation in the short-term models is a way to control for individual-specific fixed effects without having to collect longitudinal data.

PART 3:

IMPLEMENTATION STUDY DESIGN

VIII. IMPLEMENTATION STUDY DESIGN

An implementation study is an essential component of any comprehensive evaluation. A full understanding of any program requires a clear description of how that program actually operates. The findings of this implementation study have two main objectives: (1) describing the USBP and how it was implemented, and (2) understanding how the program contributed to the observed impacts.

First, by describing the program and its implementation, the study will help to better understand how the different program features are used to influence student behavior and outcomes, identify problems encountered in implementing the program, and determine how the program differs from the regular SBP. The study will also document the costs associated with program implementation and how they compare with the costs of the regular SBP.

Second, the implementation study will provide greater insight into the program's impacts by helping to determine how the program contributed to observed impacts. That is, if the pattern of impacts across related outcomes is not consistent, or if impacts vary widely by school, detailed information about program implementation will be invaluable for uncovering the underlying factors that may lead to observed impacts. Moreover, an understanding of the magnitude of differences in program features across treatment schools, or between treatment and control schools, can be useful in giving meaning to differential school impacts and to differential treatment and control group impacts.

This chapter begins by presenting the issues and topics that the implementation study will explore. It then describes how and from whom the implementation data will be collected. Finally, it lays out the plans for analyzing these data.

A. RESEARCH ISSUES

The legislation for the USBP specifies two research objectives: (1) to evaluate the effects of the USBP on student participation in the USBP and its impacts on participants' dietary intake, school achievement, and related outcomes; and (2) to determine the effect that participation in the USBP has on the paperwork and other administrative requirements placed on schools. To help address these questions, the implementation study will be guided by a set of research issues and topics that will allow the evaluation team to understand how schools have implemented the USBP, as well as what aspects of the program contributed to program impacts. Another issue to address is how meal subsidy costs change under the USBP. This chapter describes recommended research issues for the implementation study in greater detail. Seven overarching issues will provide a framework for the implementation study:

1. ***How have schools implemented the program?*** A description of how schools implemented the USBP will provide greater insight into the specific goals and objectives at each school, the different ways in which schools deal with challenges to implementation, and how well each school is working to meet its goals and objectives. To provide comparable information on how control group schools operate the SBP, a similar set of issues in the control group schools will be documented and described.

2. ***How does implementation of the program vary across schools, and to what can the evaluation attribute these cross-site differences?*** It is expected that schools will have different approaches to implementing the USBP and SBP due to different local contextual and organizational factors. The implementation study will document and analyze the reasons for these differences.

3. ***What innovative strategies have the sites implemented to create a service delivery system?*** Schools have considerable flexibility in how they actually deliver their services. A better understanding of the advantages and the challenges associated with the different modes of delivery will be helpful in identifying the most promising practices, given a school's needs.

4. ***What are the key differences between the USBP and the regular School Breakfast Program (SBP) in how they attempt to promote nutrition, learning, and other key outcomes?*** By documenting and describing a similar set of implementation issues in

both the USBP and SBP schools, the evaluation will identify key ways in which the programs differ and the ways in which these differences may influence key outcomes.

5. ***What factors about the USBP, relative to the SBP, help explain program impacts?*** An examination of both program approaches will allow the evaluation team to understand how different aspects of the program may influence student behavior and contribute to program objectives. For example, the analysis may help determine if and how the USBP helped to influence any stigma associated with consuming a school breakfast or changed student attitudes toward nutrition. In addition, information from the implementation study can allow the evaluation team to understand why the USBP is or is not differentially successful in influencing outcomes among various subgroups of students.

6. ***What are the key differences in the USBP and the SBP in terms of program costs?*** An examination of the resources required to perform different functions of the programs will allow the evaluation team to assess the relative cost-effectiveness of the USBP and the SBP. Using the findings of the USBP's impact on participation rates, coupled with data on reimbursement, the evaluation will also be able to determine the amount by which meal subsidy costs to the federal government increase under the USBP.

7. ***What changes can be made to improve the implementation and operations of the USBP?*** Based on the results of the USBP evaluation, the evaluation team will assess and identify potential program design, operations, and administrative changes that could improve the implementation of the USBP.

B. RESEARCH TOPICS

The overall plan for the implementation study incorporates a broad set of topics and data sources. To address the key issues outlined above, the implementation study will be structured around six broad research topics: (1) organizational aspects, (2) service delivery, (3) meals offered, (4) program costs and administrative burden, (5) participation, and (6) school environment. These topics will guide the development of data analysis plans, which include (1) interviews with SFA administrators, (2) interviews with school cafeteria managers, (3) interviews with school administrators and other key school staff, (4) "meals offered" data collection, (5) cost data collection, (6) staff and student focus groups, and (7) program documents and records. Table VIII.1 presents a detailed set of research topics related to various aspects of the USBP and the suggested data

TABLE VIII.1

IMPLEMENTATION STUDY TOPICS AND DATA SOURCES

	Data Sources							
	SFA Administrator	School Cafeteria Manager	School Administrator and Other Key Staff	Meals Offered Survey	Program Cost Worksheets[a]	Staff Focus Groups	Student Focus Groups	Program Documents and Records
Organizational Aspects								
Planning								
Goals and Objectives	X	X	X			X		X
Strategic Plan	X	X	X					X
Roles of the SFA and Other Agencies	X	X	X		X			X
Structure								
Key Staff Involved	X	X	X		X			X
Job Roles, Responsibilities, Level of Effort		X	X		X	X		
Training and Support Provided to Staff	X	X	X		X	X		X
Scheduling Changes to the School Day		X	X		X	X	X	
Information Systems and Available Data	X	X	X					
Service Delivery								
Recruitment and Retention of Program Participants								
Orientation		X	X		X	X	X	X
Outreach		X	X		X	X	X	X
Setting and Methods								
Breakfast Setting		X	X		X	X	X	
Delivery Methods		X	X		X			
Breakfast Schedule		X	X		X		X	
Access to Breakfast		X	X			X	X	
Promising Practices and Key Challenges	X	X	X			X	X	X

BEST COPY AVAILABLE

TABLE VIII.1 *(continued)*

	Data Sources							
	SFA Administrator	School Cafeteria Manager	School Administrator and Other Key Staff	Meals Offered Survey	Program Cost Worksheets[a]	Staff Focus Groups	Student Focus Groups	Program Documents and Records
Meals Offered								
Menu Selections		X		X				X
Nutrient Content		X		X				X
Food Provision and Usage		X		X				
Program Costs and Administrative Burden								
Program Start-up								
Program Design	X	X	X		X			X
Planning (State-, SFA-, School-Level)	X	X	X		X			X
Training	X	X	X		X			X
Outreach and Recruitment		X	X		X			X
Application Procedures		X	X		X			X
Meal Preparation								
Setting	X	X	X		X			X
Methods	X	X	X		X			X
Meal Delivery								
Setting		X	X		X			X
Methods	X	X	X		X			X
Schedule (School Day, Classes, Busing)		X	X		X			X
Reporting Requirements								
Meal Counting		X	X		X			X
Meal Content		X	X		X			X
Reimbursement								
Rates		X			X			X
Methods	X	X	X		X			X
Claiming procedures		X	X		X			X

BEST COPY AVAILABLE

TABLE VIII.1 *(continued)*

	Data Sources							
	SFA Administrator	School Cafeteria Manager	School Administrator and Other Key Staff	Meals Offered Survey	Program Cost Worksheets[a]	Staff Focus Groups	Student Focus Groups	Program Documents and Records
Participation								
Participant Satisfaction		X				X	X	
Participation Level and Patterns		X	X			X	X	X
Reasons for Participation		X	X			X	X	
School Environment								
School Profile		X	X			X	X	X
School Climate	X	X	X			X	X	
Student Behavior and Attitudes		X	X			X	X	X
Stigma		X	X			X	X	
Staff Attitudes		X	X			X	X	
Parent and School Board Support			X			X	X	

[a]It is recommended that the program cost worksheets be included as part of the interviews with the SFA administrator, school cafeteria manager, school administrator, and other school staff.

sources to address them. To understand the ways in which the USBP differs from the SBP, and to help interpret findings from the impact study, it is recommended that the evaluation collect a similar set of implementation data from the control group schools. Accordingly, the research topics and specific questions will be adapted for use with the control group schools.

1. Organizational Aspects

Organizational planning and structure can influence the implementation of any new program. If a structure is not in place to support changes, then the implementation of the program may not succeed as planned. Also, any differences in the planning and organizational structure at different sites are likely to result in variations in implementation. In order to gain a full understanding of a program, its origins must be examined. Why was the program conceived? What are the goals and objectives of the program at the local level? Who was involved in program implementation decisions, and what was their input? What measurable outcomes best express the goals of the program? An understanding of the planners' objectives and thinking in planning for the implementation of the program gives the evaluation team a blueprint for assessing the program's progress and success.

The staffing provided to implement the program can also indicate the depth of support and its potential to succeed. Which key staff are involved in operating the program, and what do they do? Were new staff hired to help implement the program? Were staff roles and responsibilities redefined as a result of the program? How? Was a reorganization of staff necessary within the school's administrative or food management offices? What is the overall level of staff effort required to implement the program?

The training, orientation, and support provided to staff are important for successful implementation of a program. Did staff undergo training and orientation? From whom and in what form do staff receive support? Has that support been sustained over time? What type of management structure or monitoring system is in place to ensure that the program runs smoothly?

Implementation of the USBP may also require changes to the school day or bus schedules in order to accommodate a larger group of students participating in the program. Changes in the length of the school day, for example, could affect the amount of effective instructional time during the school day, which in turn could explain differences in student achievement or attention span. Will schools' efforts to accommodate a larger feeding program shorten the school day? For schools that normally provide bus transportation for many of their students, there may be particular challenges associated with implementation of the USBP. How will school districts change their bus schedules? Will these changes affect schools that do not offer the USBP, including control group schools? The answers to these types of questions will allow the evaluation team to better understand how the program functions, what mechanisms are in place to support program operations, and what aspects of the program might explain differences in treatment and control group outcomes.

2. Service Delivery

The second focus of the implementation study relates to understanding how the program is operated. That is, how are services actually delivered, and how are students recruited into the program? Examining how students are recruited into the program will give the evaluation team a better understanding of participation levels and support for the program. It may also help explain, in part, differences between participation levels when the program was first implemented and over

time. For instance, was the program publicized more often when it was first implemented? What methods are used to encourage student participation in the program?

How breakfast delivery methods and settings vary between schools will also be of interest to the evaluation team. Indeed, the setting and mode of delivery chosen have important implications for a school's schedule and may also influence whether and to what extent students participate. For example, the more initiative a student must take to obtain a breakfast, the less likely the student may be to eat breakfast. Inquiring about the experiences of different schools in implementing the program, including the challenges they faced and how they dealt with them, will be helpful for identifying the most promising service delivery practices.

3. Meals Offered

Learning about the variety of foods offered as part of the USBP is essential to understanding how the program contributed to observed impacts on dietary intake and other student outcomes. The implementation study will provide a description of the prevalence of different types of foods served, the variety of foods served, and the nutrient content of those foods. The study also will examine how food is purchased and prepared, as well as how much food is being taken by students and approximately how much is being thrown away ("plate waste").

4. Program Costs and Administrative Burden

Program costs are an important consideration for the implementation study since impact estimates must be interpreted within the context of costs. Moreover, an understanding of the resources required to operationalize different program functions will allow the evaluation team to understand the relative cost-effectiveness of the USBP and SBP and to determine the effect that participation in the USBP has on program costs. It is recommended that the evaluation examine the labor and other direct

costs associated with both the initial and ongoing functions of program operation, including those related to (1) program startup, (2) meal preparation, (3) meal delivery, (4) reporting requirements, and (5) reimbursement methods and procedures. Costs associated with each of these functions, coupled with participation data, can be used to estimate a total program cost and a per meal unit cost for both the USBP and the SBP. In particular, using the findings of the USBP's impact on participation rates, the evaluation team can determine the amount by which meal subsidy costs to the federal government change under the USBP.

In examining the costs associated with implementing the various program functions, it will be important to focus on key differences between the USBP and SBP and to examine the amount of staff time and other costs associated with these differences. This will allow the evaluation team to determine which program functions are more or less expensive to operate as a result of the USBP. For example, if the number of school breakfasts served increases as a result of the USBP, there may be associated cost increases in meal preparation and delivery. That is, does the setting in which breakfast is prepared and delivered change as a result of increased participation and does this influence the number of cafeteria staff and teachers required to operate and monitor the program? Does the school schedule and/or length of the school day change as a result of the USBP? Does the level of school busing change? If so, it will be important to examine cost implications. In estimating costs, it will be particularly important to determine which staff are involved in implementing the different program functions, how much staff time is needed, and what the wage/salary and fringe benefit level are for the staff involved.

Changes in paperwork and other administrative requirements as a result of the USBP can also influence program costs. Hence, the evaluation should examine the ways in which the costs, paperwork, and administrative burden of the USBP compare to the regular SBP. For example, have

the increased costs of providing meals to more students been offset by decreases in the costs associated with the reduced handling of paperwork and other administrative tasks? How so? Have program reporting requirements and reimbursement methods been simplified through the USBP? That is, if the reporting requirements and reimbursement processes are simplified, then fewer staff resources may be required--for example, in terms of submitting claims for reimbursement--and associated program costs will decrease. In contrast, although the program eligibility and student application process will likely be simplified through the USBP, it is unlikely to have a significant effect on staff time and, in turn, program costs, since USBP schools will still be required to meet a structured set of eligibility requirements for their school lunch program. Examining these types of issues and questions will be important for understanding the overall cost-effectiveness of the USBP, relative to the SBP, and for assessing whether and how the program and its costs can be improved by making changes to specific program functions.

5. Participation

Information on participation collected through the implementation study will complement data on participant characteristics that are collected through other data sources, including the student and parent surveys and school records data. (The other data sources are described in the preceding chapters.) The implementation study will be helpful in gathering information on the overall level of student participation during a given time period, the factors that influence students to start (or stop) participating, the aspects of the program that students are most satisfied with, and the ways in which the program could be improved to better serve the target population. This information is particularly important, both for drawing policy implications and for tailoring the program to better serve the target population.

6. School Environment

Success in implementing the USBP can be affected by the environment in which it operates. Factors such as economic conditions, the labor market, the availability and quality of support services, the characteristics of the population, and the broader community's acceptance of the program may all influence the school community's attitude toward the program. For instance, if the community views a free breakfast as a handout, the stigma associated with the SBP will not be reduced by the USBP, and consequently, student impacts may be minimal. In a similar vein, the support of school and program staff, school administrators, cafeteria managers, parents, the school board, and students are all critical to the success of the program. Without the support of key players and stakeholders, the program may not meet be able to meet all of its objectives.

C. DATA COLLECTION PLANS

The research issues and questions described above will guide the development of survey instruments and interview protocols for the data collection effort. The implementation study will rely on a mix of quantitative and qualitative data collected from (1) semistructured telephone interviews with SFA administrators, (2) semistructured telephone interviews with cafeteria managers, (3) semistructured telephone interviews with school administrators and other key school staff, (4) checklists and direct observations to examine the meals offered to students, (5) worksheets to examine program costs and administrative burden, (6) focus groups with staff and students, and (7) program documents and records. The proposed sources for each of the research questions are identified in Table VIII.1; examples of specific research questions related to each topic area are outlined in Table VIII.2. Next, suggested data to collect from each of these key sources is described.

TABLE VIII.2

IMPLEMENTATION STUDY: KEY RESEARCH TOPICS AND EXAMPLES OF QUESTIONS

Research Topics	Examples of Research Questions
	Organizational Aspects
Planning	
Goals and Objectives	Why was the program conceived? What is it trying to achieve? What was the motivation for implementing the program at the local level? What is the nature of the local need for the program, and how does this need vary across sites?
	Who was involved in the program implementation decisions at both the district and the school level?
	What are the district- and school-level nutrition and food service policies?
	What measurable outcomes best express the goals of the program as it is implemented at the district and school level?
Strategic Plan	What steps were perceived as being necessary for successful implementation?
Roles of the SFA and Other Agencies	What role does the SFA play in overseeing the program?
	What other agencies are involved? What roles do they play?
Structure	
Key Staff	Was a reorganization of staff necessary within the school's administrative and food management offices?
	Were new staff hired to help implement the program?
	Who are the key staff involved in operating the program?
	Which staff are most aware of student perceptions of the program?
Job Roles and Responsibilities	Were staff's roles redefined as a result of changes brought about by the program? If so, how?
	What is the level of staff effort in implementing different aspects of the program? Has the program changed the workload of administrative staff and line staff?
	Have the roles of medical staff changed (for example, are they involved in outreach for nutrition programs)? Have they become more or less involved with students? In what way? What influence, if any, did the program seemed to have on the types of reasons students visit the nurse's office?
Training and Support Provided to Staff	Did the staff who were responsible for implementing the program undergo training or orientation?

TABLE VIII.2 *(continued)*

Research Topics	Examples of Research Questions
	From whom and in what form do staff receive support? Has that support been sustained over time?
	What type of management structure or monitoring system exists to ensure that the program runs smoothly? How does it differ from the way management was set up before the program was implemented?
Scheduling Changes to School Day	Did scheduling changes have to be made to the school day to accommodate the program?
	Did bus schedules have to change to accommodate the program?
Information Systems and Available Data	What types of information are maintained on program participants and program costs?
Service Delivery	
Recruitment and Retention of Program Participants	
Orientation	How were students and their families oriented to the new program?
Outreach	How was the program publicized?
	How, when, and how often is it publicized?
	How, when, and how often are students reminded that the program is there?
	What methods are used to encourage student participation in the program?
	What methods are used to develop parent support for the program? What methods are used to develop teacher support?
	Has outreach to students and their families changed from the time the program was first implemented to a year or two after it was implemented?
Setting and Methods	
Breakfast Setting	Where is breakfast delivered? What is the seating capacity?
Delivery Methods	How is breakfast delivered? (For example, do students get to pick what they want to eat or are they handed a package?)
Breakfast Schedule	When do students have breakfast? Were changes made to bus schedules to allow time for breakfast? Is breakfast part of the school day? How much time is allotted for breakfast?

BEST COPY AVAILABLE

TABLE VIII.2 *(continued)*

Research Topics	Examples of Research Questions
Access to Breakfast	If breakfast is not part of the school day, how much initiative do students have to take in order to eat a school breakfast? Do they, for instance, need to arrive early or skip morning recess?
Promising Practices and Key Challenges	What are the necessary steps in the process of full implementation of the program? How does this differ by site?
	What are the key characteristics on which sites vary?
	How did key program components or activities evolve or diverge from the original model?
	What factors facilitate overall program implementation and operation?
	What were the program's most significant accomplishments?
	What factors impede overall program implementation and operation? What are the biggest challenges and obstacles encountered by staff implementing the program?
	Meals Offered
Menu Selections	What is the variety of foods offered in the program? Do these selections differ from those offered before the program was implemented? How is food purchased and prepared? Is it prepared on-site or received from a vendor?
	About how large are the portions?
	What are the label ingredients on preprepared and packaged foods?
	What types of condiments are available for students' use?
	How often are hot meals served? How often are cold meals served?
	Who plans the menus? Has this changed since the program was implemented?
Nutrient Content	What is the average nutrient content of the school breakfasts offered?
Food Provision and Usage	How are foods provided (for example, as meals, a la carte, from vending machines)? How does this differ from before the program was implemented?

BEST COPY AVAILABLE

TABLE VIII.2 *(continued)*

Research Topics	Examples of Research Questions
	Program Costs and Administrative Burden
Program Startup	
Program Design	What is the staff time and what are the other resources involved in designing the program?
Planning (State-, SFA-, School-Level)	How much time was spent in planning for the implementation of the program at the state-, SFA- and school-level?
	Which staff were involved?
	What was the effect on other costs, such as travel?
Staff Training	How much time was spent training different types of staff on how to implement the program?
	Which staff were trained?
	Who provided the training?
Outreach and Recruitment	How much time was spent introducing and advertising the program?
	What other direct resources were used to promote the program?
Application Procedures	What are the paperwork and other administrative steps for a student to participate in the program? In particular, what are the eligibility determination steps for the SBP?
	How much time is required by staff?
	How do the School Lunch and other program eligibility requirements influence the time staff must spend on eligibility determination and application procedures for the SBP and USBP?
Meal Preparation	Where are meals prepared? Did the location of meal preparation change in order to accommodate an increased number of meals served? If so, how?
	Who prepares meals? Did the staff and methods of food preparation change with the new program? Are different resources used?
	How much additional staff time and other resources are required?

TABLE VIII.2 *(continued)*

Research Topics	Examples of Research Questions
Meal Delivery	Where are meals served? Did the location of meal preparation change in order to accommodate an increased number of meals served? If so, how?
	Which staff are involved in delivering meals? in monitoring student participation in the program? What are their roles? Did the staff involved and the methods of meal delivery change with the new program? If so, how?
	How much additional staff time and other resources are required?
	How did the program influence class schedules? busing schedules? How did such changes influence staff time required, for example, teacher time, cafeteria worker time, and bus driver and crossing guard time?
Reporting Requirements	What are the program reporting requirements, for example, in terms of counting the number of meals served and documenting the content or nutritional value of meals prepared?
	What steps are taken to collect the information that must be reported? to maintain the information? Which school staff perform these tasks? How are SFA- and state-level staff involved?
	How do the USBP's administrative reporting requirements compare to the SBP's? Have they been simplified? In what ways and with what implications for staff time and other costs?
	How, if at all, are costs from increased program participation offset by reduced costs and burden associated with collecting, processing, and maintaining paperwork?
Reimbursement	What are the reimbursement rates for meals served? How do reimbursement rates vary for different subgroups of students?
	What provisions are used in making claims for reimbursements? How did these change with the new program?
	What are the administrative requirements for submitting claims for reimbursement? What steps are taken in the process? What other paperwork requirements must schools meet, for example, relating to staff time and other direct costs? Which staff are involved and for how long?
	How do the USBP's paperwork and other administrative requirements relating to the reimbursement process compare to the SBP's? Have they been simplified? In what ways and with what implications for staff time and other costs?
	How, if at all, are costs from increased program participation offset by reduced costs and burden associated with collecting, processing, and maintaining paperwork?

TABLE VIII.2 *(continued)*

Research Topics	Examples of Research Questions
	Participation
Participant Satisfaction	How satisfied are participants with the program? How satisfied are the parents of participants? What are the staff perceptions of student satisfaction?
	How appealing is this program to students? In their opinions, how beneficial is the program?
Participation Level and Patterns	About how many meals were served during a given time period?
	About how many meals were consumed during a given time period?
	About how many students choose to participate on any given day?
	Are fewer students choosing to take breakfast now than when the program was first implemented and publicized?
	What factors explain differences in participation levels over time?
	Have students chosen not to participate after trying the program? Why? When did they stop participating? (That is, was it weeks, months, or years after they first participated?)
	How can the program be improved to better serve students?
Reasons for Participation	Why do students choose to participate or not participate in the program?
	What factors influence their participation decisions on a given day or in a given week?
	What role do their friends, parents, teachers, and others play in their participation decisions?
	How do participants seem to differ from nonparticipants? For example, do boys participate more than girls? Do children from relatively lower-income families participate more than those from relatively higher-income families?
	How do participants seem to differ from each other? How diverse is the pool of participants?
	School Environment
School Profile	What are the characteristics of the school and their students (for example, organizational, socioeconomic, demographic)?
	What are the characteristics of the families from which students come (for example, structure and background, labor market participation, dynamics)?
	What are the values, beliefs, and attitudes of the families from which students come (with respect to free meals, nutrition, and learning)?

BEST COPY AVAILABLE

TABLE VIII.2 *(continued)*

Research Topics	Examples of Research Questions
School Climate	What is the overall climate of the school?
	What programs and supportive services does the school provide to students?
	What problems and difficulties does the school experience?
Student Behavior and Attitudes	Do staff perceive any changes in student attitudes and behavior (toward learning or food) as a result of the program?
	Has student behavior in the classroom and at recess seemingly changed since the program was implemented? Have the number of students suspended, detained, etc. seemingly changed since the program was implemented? How has attendance seemingly changed since the program was implemented?
Stigma	Was there a stigma associated with receiving a free breakfast before the program was implemented? If so, has that changed after the program was implemented? In what way?
Staff Attitudes	How do staff feel about the program? What are their attitudes toward a USBP? What changes do they perceive as a result of the program?
	What is the level of staff turnover in the school?
Parent and School Board Support	How supportive of the program are parents and the school board?
	What is the level of parent involvement in the school?

1. Semistructured Telephone Interviews with SFA Administrators

Interviews with SFA administrators from each evaluation site will provide data on program design and implementation at the district level, especially on the planning and organizational structure, including program costs and changes in the paperwork and administrative burden. These interviews will provide valuable information on how different SFAs oversee the program and interact with other agencies with respect to the program. Moreover, the interviews will provide information on the characteristics of the SFA, its nutrition policies, and the food service operations of the schools within it.

It is suggested that the interviews be conducted over the telephone by research analysts using semistructured interviewing techniques that will allow analysts to ask questions on a basic set of topics, as well as probe for administrators' perceptions on issues relevant to program implementation in a given SFA. To ensure consistency in the data collection methods and type of information collected in each site, analysts can record answers to questions using a well-defined interview protocol that includes a mix of structured and open-ended questions. It is suggested that interviews last about 45 minutes and that they be conducted at approximately the same time as dietary intake and other student-level data are collected.

2. Semistructured Telephone Interviews with Cafeteria Managers

Telephone interviews with cafeteria managers from all the treatment and control group schools represent an important source of information on service delivery, as well as on organizational structure, school environment, cafeteria policies, and patterns of program participation. Cafeteria managers can provide information on the setting in which breakfast is delivered, the methods used to deliver breakfast, and the specific types of meals offered to students. Moreover, they will be a

useful source of information on the cost implications of the new program by providing insights on changes made to the program functions as a result of the USBP and consequent changes in the level of staff effort required. Cafeteria managers can also inform the evaluation team of any changes they observe in the cafeteria environment and in student participation patterns since implementation of the USBP.

It is suggested that telephone interviews be conducted by research analysts using semistructured interviewing techniques that will allow analysts to ask questions on a basic set of topics, as well as probe for administrators' perceptions on issues relevant to program implementation in a given SFA. Analysts should record answers to questions using a well-defined interview protocol that includes a mix of structured and open-ended questions. It is suggested that interviews last about 45 minutes and that they be conducted at approximately the same time as dietary intake and other student-level data are collected in the schools.

3. Semistructured Telephone Interviews with School Administrators and Other Key Staff

Semistructured interviews with school administrators and with other key staff from each of the treatment and control group schools will be an effective way to obtain information on planning, organizational structure, service delivery, school environment, and program participation patterns. Speaking with school administrators and other key staff will give the evaluation team a better understanding of changes that have occurred within schools as a result of the USBP. The school administrator also is likely to be the most knowledgeable source for various kinds of information on school characteristics and operations. For instance, the administrator can provide details on class size, student-to-teacher ratios, starting time and length of the school day, hours of instructional time,

changes in bus schedules, parent involvement in the school, staff turnover, availability of programs and resources (such as before-school programs), and physical education requirements.

Interviews with other key staff could yield additional information about students in the school and about the general school climate. School administrators will also be asked to identify other staff members who may be knowledgeable about program implementation and the possible influence of the program on student participation. These other staff members may include, for example, teachers, teachers' aides, nurses, guidance counselors, and cafeteria staff.

It is suggested that these interviews be conducted by research analysts using semistructured interviewing techniques that allow analysts to ask questions on a basic set of topics, as well as probe for respondents' perceptions on issues relevant to program implementation in a given school. To ensure consistency, analysts will record answers to questions using a well-defined interview protocol, which will include a mix of structured and open-ended questions. Since analysts will not conduct site visits to each school, conducting these interviews by telephone would be most cost-effective. However, when possible, analysts could conduct these interviews in person; for example, in schools where analysts conduct focus groups (described below), the school administrator and staff interviews could be conducted in person. Finally, it is suggested that the interviews be conducted over a period lasting about 45 minutes and at approximately the same time as the dietary intake and other student-level data are collected in the schools.

4. Meals Offered

A meals-offered (menu) survey will be included in the evaluation to provide descriptive information on the nutritional content of school breakfasts offered or served over the course of a week. To obtain a complete description of foods offered or served, detailed information on all foods

offered as part of a school breakfast will be needed from the school cafeteria manager. It is suggested that this information be collected by survey specialists during on-site, in-person interviews conducted in conjunction with the dietary recall data collection. In addition, vending and a la carte checklists can be completed at this time, to collect a consistent set of information across schools about the variety of foods offered or available to students, as well as the ways in which school breakfast is provided--that is, as complete meals, as single food items, and/or through vending machines.

The foods offered by the USBP and the SBP will be recorded in detail, with complete descriptions and portion sizes estimated using three-dimensional food models and, preferably, with the same automated data entry program as that used for the 24-hour-recall data collection. Similarly, the foods and amounts will be coded and analyzed for nutrient content using the same food composition database as that selected for the 24-hour recalls. When feasible, it will be useful to weigh and measure portion sizes. If several options or food bars are offered, all food items from the menus would be systematically sampled, recorded, and coded to allow for the estimation of mean nutrient availability from the school breakfast. Estimates of the percentage of meals delivered or prepared under each option could be used in concert with menu information. Food service records on food bar offerings and usage could be analyzed, along with direct observation of food bar selections. The time period covered by the meals-offered information would be either a single day (the day in which students' dietary intake information is collected) or a week-long period. Collecting data over a week would reduce the variance of the meals-offered estimates. This decrease in variance, along with cost considerations, should determine the appropriate approach. In addition, data from the meals-offered survey can be supplemented with information collected through the interviews with SFA administrators and cafeteria managers.

Verification of the accuracy of the menu information, along with information on portion sizes, would be collected from school food service personnel and/or direct observation of school breakfasts. Interviewers would observe school meals and record the foods offered, the breakfast selections available, and plate waste. Information on breakfast food selections obtained by direct observation would be entered into nutrient coding software to estimate the nutrients available from school breakfast. Observed plate waste would then be used to modify estimates of average nutrient intake available from school breakfast menus. Collection of plate waste information would require the evaluation team to obtain consent from the school and, if individual children are observed, from parents. (The issue of obtaining consent is discussed in more detail in Chapter VI.) Observing individual children's intake may provide an external verification of reported breakfast intake, but it would not necessarily provide better information than reporting the breakfast intake immediately after breakfast during the dietary data collection.

5. Program Costs

To develop an estimate of program costs, the evaluation team will rely on an approach that bases cost estimates on information collected through staff interviews about the use of staff and other resources in implementing the program (Ohls and Rosenberg 1999). Although the evaluation team should also obtain administrative cost data from the schools and SFAs, such program data are unlikely to provide sufficient detail on specific program functions or to be reported in a consistent format across schools, thus making it difficult to rely on administrative data to estimate program costs. Hence, it is suggested that the evaluation team develop and use structured cost data

worksheets to collect information from program staff that will allow them to estimate the different types of costs that are incurred in operating the USBP and SBP programs.

Using the "building-up" cost estimation approach developed in Ohls and Rosenberg (1999), which has been used effectively in other social program evaluations, the evaluation team can first collect descriptive information on the resources used--most importantly, the staff time used--in implementing each of the key program functions described above. Then the evaluation team can develop program cost estimates based on the level of resources used and relevant unit costs, focusing on those program functions where the USBP and SBP differ. Since this approach requires interviewing program administrators and line staff, the cost data worksheets can be administered as part of the interviews with the SFA administrator, the school cafeteria manager, school principal, and other school staff (particularly those associated with the cafeteria). The majority of questions will be posed to the cafeteria manager, two or three key cafeteria staff, and the school principal, who are likely to have the most information on how the program is implemented.

Questions included in the cost data worksheets should be designed to measure how much staff time is required to perform different program functions. By categorizing the types of staff involved in implementing specific program functions, estimating the amount of time they spend each week (or each month), and identifying their wage/salary and fringe benefit rate, the evaluation team can develop labor cost estimates for each program function. Labor cost estimates are most critical to the analysis since they represent the majority of program costs. Information on other direct costs--such as food and nonfood supplies, rent and space, storage, utilities, insurance, and equipment--is typically available in some form from program records and can be used to complete the program cost estimates, either by attributing these costs to specific program functions or by estimating them in

proportion to labor and allocating them to particular functions as appropriate. The value of donated goods and services should also be considered part of the analysis.

It is suggested that the cost data interviews and worksheets be completed over the telephone by research analysts using cost data worksheets that are comprised of structured and detailed sets of closed-end questions. Worksheets can be included in the survey interview protocols for the program staff described above. As necessary, cost-focused telephone interviews of 10 to 15 minutes may need to be completed with one or more additional cafeteria staff to round out the cost data collection. To maximize the usefulness of the telephone interview process, analysts can begin with an initial telephone contact, follow up with a letter to the interview respondent explaining the purpose of the data collection effort, and then conduct the interview.

6. Focus Groups

Focus groups--small group discussions with 6 to 10 individuals--will be valuable for gathering insights from staff and students on the factors that may explain:

- Student participation in the program
- The level of student, parent, and school support for the program
- The nature of the school environment
- The service delivery practices and methods that are most effective
- The ways in which the program could be improved to maximize student participation

The evaluation team will conduct informal focus groups in selected treatment and control group schools with groups of both staff and students. Focus groups can be conducted by research analysts and survey specialists experienced in focus group moderation. The discussions should be guided

by structured protocols and procedures designed to encourage participants to talk candidly. It is suggested that they be conducted during the same time period that the dietary intake and other student-level data are collected in the schools.

Through discussions with school administrators, the evaluation team can identify the best strategy for recruiting staff and students to participate in the focus groups. Staff focus groups might include a mix of different types of staff who are familiar with the program and with student participation in it. For example, a staff focus group might include a mix of teachers, teachers' aides, nurses, guidance counselors, and cafeteria staff. Student focus groups should include only students in the higher grades, such as the fifth and sixth grades. Separate student focus groups might be considered for boys and girls in order to minimize inhibitions about freely responding to questions.

7. Program Documents and Records

Documents and records about the program and about the school context will be requested from the SFAs and schools to supplement the data collected through interviews, the meals offered component, and focus groups. These documents and records will include, for example, school food service records, administrative cost data, descriptive information about the study schools, and written materials documenting the program and the steps the SFA and schools took to implement it. Training materials and memos provided to staff, as well as materials about the program that were given to parents and students, could be requested as well. In addition to these documents, school records data--such as aggregate data on student characteristics like attendance, achievement, and disciplinary incidents--should also be examined. These school records data will be acquired as part of the data collection plan for the impact evaluation (described in an earlier chapter).

D. DATA ANALYSIS AND SYNTHESIS

A multifaceted approach to the analysis and interpretation of data on program implementation will be employed to produce a comprehensive description of the implementation of the USBP in the study schools. This approach involves both quantitative and qualitative analysis of the data summarizing the six key aspects of the program: (1) the organizational structure supporting the USBP, (2) the delivery of program services, (3) the meals offered to students, (4) program costs and administrative burden, (5) the factors explaining participation in the program, and (6) the school environment in which the program operates. Implementation analysis presents the challenge of combining information from various sources in a systematic approach to analysis and inference. Thus, it is suggested that two guiding principles be followed:

1. ***Create a structure for analysis of the data in advance.*** All the information collected for the implementation analysis must be analyzed in a consistent framework. Therefore, it is important that an analytic framework for the analysis be developed. The conceptual framework introduced in Chapter II and the research topics and questions outlined above represent the starting point for the development of such a framework.

2. ***Triangulate sources and perspectives.*** It is critical that data collected through different strategies and from different sources be analyzed as a whole. Thus, the plan for analyzing the data includes identification of all sources from which data are to be collected on similar issues, as well as identification of ways in which data from multiple sources can be used to validate findings.

In general, the implementation analysis will proceed "from the inside out." In other words, it will focus first on the services and the modes of delivery, then on how external factors and program organization affect services. Next, the analysis will assess, from a broad perspective, how the overall design and structure of the USBP affects implementation at the school level. Finally, it will assess how these factors and influences might affect the quality of service implementation and the outcomes achieved. Ultimately, the analysis will generate lessons and recommendations for program design at the school level and policy decisions at the district and national levels.

1. Analysis of Quantitative Data

The quantitative analysis will examine data on the structure of the USBP across study schools, along with the different methods used to recruit students to the program, the service delivery settings and methods used, the types of meals offered to students (including their nutrient content), program costs, and the characteristics of the school environment in which the program is implemented. Similar descriptive data will be examined for the control group schools and their SBPs. The analysis will consist of examining (1) descriptive statistics, using schools as the unit of analysis to compute unidimensional summary measures such as averages, frequencies, and percentiles; and (2) multidimensional cross-tabulation measures to help assess variation in the characteristics of USBP implementation across schools. Similar analyses will examine patterns across control schools and between treatment and control schools. A broad, descriptive profile of how the USBP is implemented in study schools can be developed using descriptive data from the interviews and the meals-offered data collection. This analysis will be enriched by qualitative data on program implementation collected through the interviews and focus groups.

2. Analysis of Qualitative Data

The qualitative analysis of implementation data includes both descriptive and explanatory methods both within individual schools and across schools. Qualitative analysis *within schools* will be descriptive and will provide some assessment of program implementation. Interview guides and protocols can be used to structure the data collection for the analysis of information gathered during interviews and focus groups. Qualitative descriptive information can be synthesized on each treatment school for information about, and preliminary analysis of, how each school is implementing the USBP.

Qualitative findings can also be synthesized and assessed *across schools*. That is, experiences will be compared across schools to identify similarities and differences in the implementation and operation of the USBP. These data also will help explain what led the schools to use different practices in implementing the program. Tabular summaries of key program features can be created to assist in the cross-school analysis. From these analyses, lessons will be drawn about program implementation for individual schools, as well as for school districts. Lessons may involve, for example, the identification of both promising practices and critical challenges related to implementing the program. The analysis of cross-school differences will take into account the different school contexts and target populations. Similar analyses should be conducted for control group schools and their SBPs. The evaluation team could also aim to identify important implementation differences between the USBP and SBP schools.

3. Integration of Quantitative and Qualitative Data

Quantitative and qualitative data can be combined to present a comprehensive analysis of how the USBP has been implemented in various SFAs and schools. Each type of data presents a different view of implementation. When combined, these two types of data will provide as complete a picture as possible of the implementation of the USBP in different school contexts. The quantitative data will provide a snapshot of how the day-to-day operation of the USBP translates into program delivery and the meals offered to students. The qualitative data will provide a detailed picture of how the USBP has evolved over time, including, for example, the different methods used to deliver services and the factors that may help explain student participation in the program and key differences between participants and nonparticipants. Together, the two types of data will provide a more complete picture of program implementation than either can provide alone.

PART 4:

FEASIBILITY OF IMPLEMENTATION

IX. OVERALL FEASIBILITY OF EVALUATION DESIGN OPTIONS

This report has developed four options for evaluating the effects of the Universal-Free School Breakfast Program (USBP) on student outcomes. All the designs are based on an approach that pairs schools within districts and then randomizes each school in the pair into either the USBP or regular School Breakfast Program (SBP) groups. The designs differ in terms of (1) whether baseline and follow-up data on student achievement on sampled students will be provided by the school districts, and (2) whether the evaluation will include a preimplementation survey of students' parents. The choice of the final design can only be made after FNS finds out which school districts apply for the USBP demonstration.

This chapter assesses the overall feasibility of these designs options. Several dimensions are considered, including the ability of the evaluation to detect program impacts, demonstration and evaluation costs, the risks associated with implementing the design, and the potential for successful completion. Table IX.1 provides a summary of the design options.

A. OVERVIEW OF THE DESIGN OPTIONS

In this section, the design options are first described. Note that, under each design, school records data will be collected. These records include school-level summary variables (school means), based on all students attending schools in the study, and, where appropriate, data on sampled students. School records data will cover four school years: the three school years in which the demonstration is active (school year 2000-2001, school year 2001-2002, and school year 2002-2003, and one year prior to demonstration startup (school year 1999-2000).

TABLE IX.1

SUMMARY OF UNIVERSAL-FREE SCHOOL BREAKFAST PROGRAM IMPACT EVALUATION DESIGN OPTIONS

	Design Alternative 1: Baseline and Followup District-Administered Achievement Data Available		Design Alternative 2: Evaluation Contractor Must Conduct a Follow-Up Achievement Test	
	(1a) Without Preimplementation Survey[a]	(1b) With Preimplementation Survey[a,c]	(2a) Without Preimplementation Survey[a]	(2b) With Preimplementation Survey[a,c]
Approach	Pair and randomize schools Regular SBP is the counterfactual	Pair and randomize schools Regular SBP is the counterfactual	Pair and randomize schools Regular SBP is the counterfactual	Pair and randomize schools Regular SBP is the counterfactual
Sample Size[b]	144 schools; 4,320 students	144 schools; 4,320 students	120 schools; 3,600 students	120 schools; 3,600 students
MDEs: Experimental-Based Methods				
Participation (in percentage points)	6.7	6.7	7.3	7.3
Dietary Intake (percentage of RDA)				
Food energy	6.0	3.5	6.5	3.9
Calcium	11.4	6.6	12.5	7.3
Academic Achievement (percent of standard deviation)	27.2	16.0	29.8	18.0
MDEs: Nonexperimental-Based Methods				
Dietary Intake (percentage of RDA)				
Food energy	1.9	1.9	2.1	2.1
Calcium	3.6	3.6	3.9	3.9
Academic Achievement (percent of standard deviation)	8.7	8.6	9.5	9.4
Demonstration Costs[d]	$6.7 million		$5.6 million	
Limitations/Risks	Can reliably estimate impacts on some but not all dietary outcomes using experimental based analysis methods without preimplementation survey, but cannot reliably estimate impacts on student achievement. If preimplementation survey is conducted and reasonably accurate, then would be able to reliably detect impacts on dietary outcomes but not student achievement, using experimental-based analysis methods. Fallback position is to estimate impacts using nonexperimental methods by comparing regression-adjusted mean outcomes of new USBP participants and nonparticipants from SBP schools. However, if selection bias important--cannot reliably estimate impacts on student achievement.		Cannot reliably estimate impacts on dietary outcomes or student achievement using experimental based analysis methods without preimplementation survey. If preimplementation survey is conducted and reasonably accurate, then would be able to reliably detect impacts on dietary outcomes but not student achievement, using experimental-based analysis methods. Fallback position is to estimate impacts using nonexperimental methods by comparing regression-adjusted mean outcomes of new USBP participants and nonparticipants from SBP schools. However, if selection bias important--cannot reliably estimate impacts on dietary intake and achievement.	

TABLE IX.1 *(continued)*

NOTES: See tables in Section V.C.4 for formulas used to calculate precision of estimates.

[a] If the demonstration startup is delayed to January 2001, then the evaluation will include a preimplementation survey to attempt to identify students in the treatment and control groups most likely to become new breakfast program participants. Precision estimates shown assume preimplementation survey predicts new participants at least 50 percent of the time (that is, of the parents who are surveyed say that their student would become a new participant and be compared in the treatment and control groups, 50 percent of the students would, in fact, become new participants under the USBP).

[b] The design is a "balanced" design, in which half of the school s are USBP schools and half regular SBP schools. Demonstration costs will be lower if an "unbalanced" design, that is, one with more SBP than USBP schools. Statistical precision will be somewhat lower under an unbalanced design.

[c] Calculations assume the preimplementation survey method will enable the evaluation to accurately predict "would be new participants" at least 50 percent of the time (that is, of the students currently not participating in the regular SBP who are surveyed that say they would participate in the USBP and being compared in the treatment and control groups, 50 percent of them would in fact become new participants under the USBP).

[d] Demonstration costs refer to meal subsidy costs for USBP schools. The estimates assume participation will increase 25 percentage points in USBP schools (from 30 to 55 percent). Evaluation costs are not included. However, preliminary estimates indicate that all the design options fit approximately within the $13 million available for demonstration and evaluation costs.

MDE = minimum detectable effect

1. Design 1a: Randomize 144 Schools, No Preimplementation Survey

Starting with 144 schools, Design 1a would first pair schools and then randomly assign each of the schools in each pair to either a treatment (USBP) or a control (regular SBP) group, resulting in 72 schools in each group. Since the design does not include a preimplementation survey, the demonstration can start in fall 2000. A sample of 30 students would be selected from each demonstration school, for a total sample of 4,320 students.

Under Design 1a, the evaluation contractor would, in spring 2001, administer one round of the full set of SFA, principal, cafeteria manager, teacher, parent, and student surveys (the student survey includes dietary intake recalls and cognition tests). Data on student academic achievement test measures would be obtained from school records, covering at least two years (baseline and follow-up test data). The evaluation would also collect school records on other outcomes covering four years, from fall 1999 through spring 2003.

Two partially overlapping samples would be included in Design 1a, a cross-sectional sample and a longitudinal sample. The cross-sectional sample would include students in grades 1 through 6 in spring 2001. Members of the cross-sectional sample would be administered all questionnaires. The panel component consists of a sample of students in grades 1 through 5 in spring 2001 (grades 2 through 5, if three years of achievement data are available).

The evaluation would submit an interim report in spring 2002. Under Design 1a, the evaluation's interim report would report on USBP impacts on participation and all other outcomes based on student-level data, including student achievement. The report would also include analyses of data available from administrative school records, such as breakfast program participation and attendance. The analyses of student records data for the interim report would cover school year 1999-2000 (the preimplementation period) through school year 2000-2001.

The final report would summarize longer-term findings on the impacts on student outcomes using student-level records data--such as participation, achievement, attendance, and discipline--covering the period school year 1999-2000 through school year 2002-2003.

2. Design 1b: Randomize 144 Schools, Conduct Preimplementation Survey

Design 1b is the same as Design 1a, except that it includes a preimplementation survey of students' parents in the fall 2000. The survey would be conducted only if the demonstration start-up is delayed until January 2001. Prior to actual USBP implementation, the evaluation contractor would select a sample of 100 students from each school (both USBP and regular SBP schools) participating in the demonstration. The sampled students (and/or parents) would be administered a brief telephone survey, the objective of which is to enable the evaluation contractor to identify students most likely to be new USBP participants. The survey would ask about the student's usual participation in the regular SBP, and whether the student would participate in a universal-free school breakfast program. This survey would be administered in fall 2000.[1] A sample of 30 students would then be selected from the 100 student sample from each demonstration school. Students currently not participating in the SBP, but who indicate they would participate in a USBP, would be oversampled.

As with Design 1a, the evaluation contractor under Design 1b would administer one round of the full set of surveys in spring 2001; data on student achievement would be collected from

---

[1]If possible, it is desirable to administer this survey in summer 2000 and implement the demonstration in fall 2000. This would require having the evaluation contract in place by summer 2000 and securing early cooperation of the districts to obtain lists of students, in order to select the samples.

administrative records.[2] The evaluation would collect other school records data covering four years. Analysis and reporting for the interim and final report is the same under Design 1b as those described for Design 1a.

3. Design 2a: Randomize 120 Schools, No Preimplementation Survey

Design 2a is the same as Design 1a, except that the evaluation contractor would need to conduct a single round of follow-up tests of student academic achievement for members of the longitudinal sample. (The baseline measure of achievement would come from student records data.) Having to spend resources to administer and process the achievement tests means that fewer schools and students could be sampled. Design 2a therefore includes 120 schools--60 USBP schools and 60 regular SBP schools--and a total of 3,600 students. However, the schedule for data collection, analysis, and reporting is the same as that for Design 1a.

4. Design 2b: Randomize 120 Schools, Conduct Preimplementation Survey

Design 2b is the same as Design 2a, except that it includes a preimplementation survey. Again, the survey would be conducted only if the demonstration startup is delayed. The survey would be administered in fall 2000, with demonstration startup occurring in January 2001. As with Design 2a, the evaluation contractor under Design 2b would administer a follow-up achievement test; the baseline measure of student achievement would be obtained from school records.

---

[2]With a January 2001 startup, conducting student surveys and the other surveys in spring 2001 may not give students enough time to adjust their participation to the new meal program. It may be necessary to delay the full set of student, teacher, and staff surveys until the following spring. If the survey data collection is delayed until spring 2002, then the interim report would include only findings from administrative records data; findings based on analysis of survey data and longer-run findings from administrative data analysis would be available in the final report.

B. PRECISION OF THE IMPACT ESTIMATES

This section contrasts the four design options on the precision in which they would detect USBP impacts on student outcomes. Four outcomes are considered--participation in the school breakfast program, intake of food energy at breakfast, intake of calcium at breakfast, and student achievement in reading.

1. Participation in the USBP

All the designs would be able to reliably detect the expected size of program effect in the ranges of 7.3 to 8.3 percentage points on participation using school-level data, and 6.7 to 7.3 percentage points using student-level data (see Table IX.1). Thus, any of the design options will easily be capable of detecting the target minimum detectable difference of 25 percentage points.

2. Dietary Intake at Breakfast

Design 1a would be able to reliably detect some, but not all, of the target detectable differences for dietary intake using experimental-based analysis methods; and Designs 1b and 2b would be able to do so as long as the preimplementation survey method was reasonably accurate (see Table IX.1). Using experimental-based analysis methods, Design 1a, with 144 schools and 4,320 students, would be able to detect differences in food energy expected under the USBP, but not calcium. Although close to the target objectives, Design 2a, with 120 schools and 3,600 students, would not be able to detect differences in either food energy or calcium. Both designs (Design 1b and 2b) would be able to reliably detect impacts on dietary outcomes if a preimplementation survey were conducted and the survey was reasonably accurate. Much more acceptable levels of precision could be achieved for all four designs using nonexperimental analysis methods if selection bias problems proved not to be serious.

3. Student Academic Achievement

None of the designs would reliably detect differences in student academic achievement using experiment-based analysis methods (see Table IX.1). If successful, the strategy based on conducting a preimplementation survey yields more precise estimates of USBP impacts on student achievement than could be achieved without the survey. However, the minimum detectable estimates attained with a preimplementation survey would still be well above what could reasonably be expected to occur under the USBP.

Design 1a would be able to detect an impact on student achievement as low as 8.7 percentage points (3 percentiles) using nonexperimental-based analysis methods, so long as selection bias was not a problem. Precision could be improved only slightly--to 8.6--if a preimplementation survey were also part of the design (see Design 1b) and it was reasonably accurate. Design 2a would be able to detect a 9.5 percentage point change in student achievement using nonexperimental methods, and 9.4 percentage point change if an accurate preimplementation survey were included in the design (Design 2b).

C. DEMONSTRATION AND EVALUATION COSTS

Approximate costs of the demonstration and evaluation have been estimated for the four designs under consideration. (These estimates should be considered preliminary, and are subject to refinement once more information is available about the SFAs participating in the program.) All the designs fit approximately within the $13 million funding constraint. Switching to an unbalanced design, where more schools are assigned to SBP than USBP schools, could lower demonstration costs somewhat, if sufficient schools are available; other adjustments in data collection are possible as well.

D. POTENTIAL FOR SUCCESSFUL COMPLETION

Each of the designs has limitations and risks, which are discussed below. An important fallback position for all the designs is to directly estimate impacts on new participants in the USBP by comparing regression-adjusted mean outcomes of new USBP participants with nonparticipants from regular SBP schools. If selection bias is not important, the evaluation will be able to detect impacts on academic achievement. If selection bias is important, however, the evaluation cannot reliably estimate impacts on student achievement.

1. Risks of Implementing Design 1a and Design 2a

Designs 1a and 2a probably are the least risky of the alternative research strategies. However, as discussed earlier, because the treatment versus control status in these designs are clustered by school, it is not possible to reliably detect differences in many of the dietary intake outcomes nor school achievement of the sizes likely to occur under the demonstration, using experimental-based analysis methods. For example, in its basic form, Designs 1a and 2a will reliably detect only an extremely large effect on student achievement--on the order of 11 to 12 percentile points. Should selection bias problems not be a serious problem, acceptable levels of precision could be achieved for these designs using nonexperimental analysis methods.

2. Risks of Implementing Design 1b and Design 2b

Administering a preimplementation survey of students' parents and using the data to identify students most likely to be new participants could increase the power of the designs. If the approach accurately identified new participants, the evaluation could detect changes of approximately four percentage points in food energy and seven percentage points in calcium using experimental-based analyses, which is well within the target range. However, these designs would not reliably detect

impacts on student achievement, using experimental-based analysis methods. Nonexperimental analysis methods would still be required to reliably estimate target-level impacts on student achievement. Of course, all of this is conditional on selection bias not being a serious problem in the nonexperimental analyses.

It may be difficult in practice to accurately predict the students in treatment and control groups most likely to be new participants. Parents will be asked whether their children would participate in the USBP. These responses, and other student and parent characteristics data, will be used to predict the students most likely to become new participants. Previous evaluations that rely on similar approaches have met with mixed success. There is the additional complication that, if the demonstration is not delayed, then the evaluation cannot include a preimplementation survey. The approach, although risky, would have a large payoff if successful; it is worth the modest resources needed to implement it.

E. RECOMMENDED APPROACH

Any of the four options, and either of the possible analysis approaches, represents a reasonable research strategy. The option ultimately selected will depend on the SFAs applying for the demonstration. If a large number of SFAs apply and many conduct student achievement tests on an annual basis, then FNS can select districts that could provide achievement test data. In that case, either Design 1a or 1b is the preferred approach. Further, if feasible, it makes sense to conduct a preimplementation survey and oversample students based on that survey. Oversampling allows the evaluation to increase the power of the student outcome analysis without unduly compromising the ability to analyze other outcomes and intermediate variables. The major drawback of a preimplementation survey is the possibility that such a survey would delay USBP implementation. For that reason, Design 1b would be selected over Design 1a only if the demonstration were to be

delayed for other reasons. If the demonstration is not delayed, then Design 1a (or Design 2a) is the preferred option.

REFERENCES

Abell Foundation. "Data from Abell Foundation Project Concludes: Changes in Student Breakfast Program Will Increase Participation, Improve Performance." The Abell Report, 1998, vol. 11, pp. 1-5.

Blumberg, S.J., K. Bialostosky W.L. Hamilton, and R.R. Briefel. "The Effectiveness of a Short Form of the Household Food Security Scale." *American Journal of Public Health*, vol. 89, 1999, pp. 1231-1234.

Briefel, Ronette, J. Michael Murphy, Susanna Kung, and Barbara Devaney. "Breakfast and Learning: A Review of the Literature." Draft report prepared for the U.S. Department of Agriculture, Food and Nutrition Service. Princeton, NJ: Mathematica Policy Research, Inc. October 1999.

Brown, J. Larry, and Pollitt Ernesto. "Malnutrition, Poverty, and Intellectual Development." *Scientific American*, February 1996.

Bryk, Anthony S. and Stephen W. Raudenbush (1992) Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park, CA: Sage Publications.

Burghardt, John, Anne Gordon, Nancy Chapman, Philip Gleason, and Thomas Fraker. "The School Nutrition Dietary Assessment Study: School Food Service, Meals Offered, and Dietary Intakes." Alexandria, VA: USDA Food and Nutrition Service, October 1993a.

Burghardt, John, Todd Ensor, Gayle Hutchinson, Charlene Weiss, and Bruce Spencer. "The School Nutrition Dietary Assessment Study: Data Collection and Sampling." Alexandria, VA: USDA Food and Nutrition Service, October 1993b.

Burghardt, John, Anne Gordon, and Thomas Fraker. "Meals Offered in the National School Lunch Program and the School Breakfast Program." *American Journal of Clinical Nutrition*, vol. 61(suppl), 1995, pp. 187S-198S.

Carriquiry, Alicia L. Assessing the prevalence of nutrient inadequacy. *Public Health Nutrition* 1998; vol. 2, no. 1, pp 23-33.

Center for Applied Research and Educational Improvement. *Minnesota Universal Breakfast Pilot Study: Year Three Report, 1996-97.* CAREI, University of Minnesota, 1997.

Center for Applied Research and Educational Improvement. *Minnesota Universal Breakfast Pilot Study: Two-Year Report, 1995-96.* CAREI, 1996.

Center on Hunger, Poverty and Nutrition Policy. *The Link Between Nutrition and Cognitive Development in Children.* Tufts University School of Nutrition, 1994.

Chandler, A.K., S.P. Walker, K. Connolly, and K. Grantham-McGregor. "School Breakfast Improves Verbal Fluency in Undernourished Jamaican Children." *Journal of Nutrition*, vol. 125, April 1995, pp. 894-900.

Coleman, James S., Thomas Hoffer, and Sally Kilgore. *High School Achievement: Public, Catholic, and Private Schools Compared.* New York: Basic Books, 1982.

Cook, J.T., O. Punam, and G.L. Kelly. "Evaluation of a Universally-Free School Breakfast Program: Central Falls, Rhode Island." Center on Hunger, Poverty, and Nutrition Policy, Tufts University School of Nutrition Science and Policy, January, 1996.

Dairy Council Digest. "Breakfast: Its Effects on Health and Behavior." March 1993.

Devaney, Barbara, and Elizabeth Stuart. "Eating Breakfast: Effects of the School Breakfast Program." Final report to the Food and Nutrition Service. Princeton, NJ: Mathematica Policy Research, Inc., July 13, 1998.

Devaney, Barbara, Anne Gordon, and John Burghardt. "Dietary Intakes of Students." *American Journal of Clinical Nutrition*, vol. 61(suppl), 1995, pp. 205S-212S.

Devaney, Barbara, and Thomas Fraker. "The Dietary Impacts of the School Breakfast Program." *American Journal of Agricultural Economics*, vol. 71, no. 4, November 1989.

Ervin, RB, J.D. Wright, and J. Kennedy-Stephenson. "Use of Dietary Supplements in the United States, 1988-94. National Center for Health Statistics." *Vital and Health Statistics*, vol. 11, no. 244, 1999. Hyattsville, MD: National Center for Health Statistics.

Glazerman, Steven M., and Barbara Devaney. "Extended Measures of Diet Quality: Options for Studying the Adequacy of the Thrifty Food Plan." Draft report prepared for the USDA Food and Nutrition Service. Princeton, NJ: Mathematica Policy Research, Inc, 1998.

Goldhaber, Dan D., and Dominic J. Brewer. "Why Don't Schools and Teachers Seem to Matter?" *Journal of Human Resources*, vol. 32, no. 3, 1997, pp. 505-523.

Greene, William H. (1993). Econometric Analysis, 2nd Edition. New York: MacMillan. pp. 280ff.

Hack, Maureen, Nancy K. Klein, and H. Gerry Taylor. "Long-Term Developmental Outcomes of Low Birth Weight Infants." *The Future of Children*, vol. 5, no. 1, spring 1995, pp. 176-196.

Hanushek, Erik. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*, vol. 24, no. 3, 1986, pp. 1141-1177

Jacoby, E., S. Cueto, and E. Pollitt. "Benefits of a School Breakfast Program Among Andean Children in Huartz, Peru." *Food Nutrition Bulletin*, 1996, vol. 17, pp. 54-64.

Jacoby, E., S. Cueto, and E. Pollitt (in press). "Breakfast Prevents Delays of Attention and Memory Functions Among Growth-Retarded Children." *Journal of Applied Developmental Psychology*.

Kennedy, Eileen, and Carole Davis. "U.S. Department of Agriculture School Breakfast Program." *American Journal of Clinical Nutrition*, vol. 67(suppl), 1998, pp. 798S-803S.

Kleinman, R.E., J. M. Murphy, M. Little, C.A. Wehler, K. Regal, and M.S. Jellinek. "Hunger and Psychosocial Functioning: Potential Behavioral and Emotional Correlates." *Pediatrics*, 1998, 101:e3.

Kovacs, M. "The Children's Depression Inventory (CDI)." *Psychopharmacology Bulletin*, 1985, vol. 21, pp. 995-998.

Landgraf, J.M. Abetz, L., and J.E. Ware. *Child Health Questionnaire (CHQ): A Users Manual*. First edition. Boston: The Health Institute, New England Medical Center, 1996.

Levitsky, David. "Malnutrition and the Hunger to Learn." In *Malnutrition, Environment, and Behavior*, edited by D.A. Levitsky. Ithaca, NY: Cornell University Press, 1979, pp. 161-179.

Marcotte, Lori P. "A Guide to Universal Free School Breakfast Programs." Medford, MA: Center on Hunger and Poverty, School of Nutrition Science and Policy, Tufts University, 1999.

Meeks, Gardner, Julie M., Sally M. Grantham-McGregor, Susan M. Chang, John H. Himes, and Christine A. Powell. "Activity and Behavioral Development in Stunted and Nonstunted Children and Response to Nutritional Supplementation." *Child Development*, vol. 66, 1995, pp. 1785-1797.

Meyers, A.F., A.E. Sampson, M. Weitzman, and B.L. Rogers. "School Breakfast Program and School Performance." *American Journal of Diseases in Children*, vol. 143, 1989, pp. 1234-1239.

Miller, J.S., and S. Korenman. "Poverty and Children's Nutritional Status in the United States." *American Journal of Epidemiology*, vol. 140, no. 3, 1994, pp. 233-243.

Minnesota State Department of Children, Families, and Learning. *School Breakfast Programs: Energizing the Classroom*. 1996.

Mirsky, A.F., B.J. Anthony, C.C. Duncan, M.B. Ahearn, and S.G. Kellam. "Analysis of the Elements of Attention: A Neuropsychological Approach." *Neuropsychological Review*, vol. 2, 1991, pp. 109-145.

Mosteller, Frederick. "The Tennessee Study of Class Size in the Early School Grades." In *The Future of Children*, vol. 5, no. 2, Summer/Fall 1995. www.futureofchildren.org

Murphy, J.M., I.M. Kopits, R. Hicks, J. Marinaccio, and R.E. Kleinman. "Effects of a Universally Free, In-Classroom School Breakfast Program: Results from the Maryland Meals for Achievement Study." 1999. Unpublished manuscript, Massachusetts General Hospital.

Murphy, J.M., M. E. Pagano, R.E. Kleinman, and M.S. Jellinek. "The Relationship Between Hunger and Psychosocial Functioning in Low-Income American Children." *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 37, 1998(a), pp. 163-170.

Murphy, J.M., M. Pagano, J. Nachmani, S. Kane, M. Little, and R. Kleinman. "The Relationship Between School Breakfast and Psychosocial and Academic Functioning: Results from an Inner City Sample." *Archives of Pediatrics and Adolescent Medicine*, vol. 152, 1998(b), pp. 899-907.

National Research Council. *Diet and Health: Implications for Reducing Chronic Disease Risk*. Washington, DC: National Academy Press, 1989a.

Nicklas, Theresa, Carol E. O'Neil, and Gerald S. Berenson. "Nutrient Contribution of Breakfast, Secular Trends, and the Role of Ready-to-Eat Cereals: A Review of Data from the Bogalusa Heart Study." *American Journal of Clinical Nutrition*, 1998 (supplement), p. 67.

Nicklas, Theresa A., et al. "Breakfast Consumption Affects Adequacy of Total Daily Intake in Children." *Journal of the American Dietetic Association*, vol. 93, no. 8, August 1993.

Nusser, S. M., A.L. Carriquiry, K.W. Dodd, and W.A. Fuller. "A Semiparametric Transformation Approach to Estimating Usual Dietary Intake Distributions." *Journal of the American Statistical Association*, vol. 91, 1996, pp. 1440-1449.

Office of Technology Assessment, U.S. Congress. *Testing in American Schools: Asking the Right Questions*. Washington, DC: U.S. Government Printing Office, February 1992.

Ohls, J.C. and L.C. Rosenberg. "A "Building-Up" Approach to Measuring Program Costs." *Journal of Policy Analysis and Management*, vol. 18, no.3, 1999, pp. 473-480.

Pollitt, E. "Breakfast, Cognition, and School Learning." Supplement to the *American Journal of Clinical Nutrition*, vol. 67, 1998 (supplement), pp. 746S.

Pollitt, Ernesto, and Rebecca Mathews. "Breakfast and Cognition: An Integrative Summary." *American Journal of Clinical Nutrition*, vol. 67 (supplement), 1998.

Pollitt, Ernesto, Santiago Cueto, and Enrique R. Jacoby. "Fasting and Cognition in Well- and Undernourished Schoolchildren: A Review of Three Experimental Studies." *American Journal of Clinical Nutrition*, vol. 67 (supplement), 1998.

Pollitt, Ernesto, Mari Golub, Kathleen Gorman, Sally Grantham-McGregor, David Levitsky, Beat Schurch, Barbara Strupp, and Theodore Wachs. "A Reconceptualization of the Effects of Undernutrition on Children's Biological, Psychosocial, and Behavioral Development." *Social Policy Report*, vol. 10, no. 5, 1996, pp. 1-22.

Pollitt, E. "Does Breakfast Make a Difference in School?" *Journal of the American Dietetic Association*, vol. 95, 1995, pp. 1134-1139.

Pollitt, Ernesto, and Rudolph L. Leibel. "Iron Deficiency and Behavior." *Journal of Pediatrics*, vol. 88, no. 3, 1976, pp. 372-381.

Ponza, Michael, Dexter Chu, Mary Kay Fox, and William Rhodes. "Analysis of Nutrition Outcomes of Students." Final report to the Food and Nutrition Service. Princeton, NJ: Mathematica Policy Research, Inc., February 28, 1997.

Powell, C.A., S.P. Walker, S.M. Chang, and S.M. Grantham-McGregor. "Nutrition and Education: A Randomized Trial of the Effects of Breakfast in Rural Primary School Children." *American Journal of Clinical Nutrition*, vol. 68, 1998, pp. 873-879.

Psychological Assessment Resources, Inc. (1999). Online Catalog of Testing Resources. www.parinc.com

Reynolds, C.R., and B.O. Richman. *Revised Children's Manifest Anxiety Scale (RCMAS)*. Los Angeles, CA: Western Psychological Services, 1985.

Sampson, Amy, et al. "The Nutritional Impact of Breakfast Consumption on the Diets of Inner-City African-American Elementary School Children." *Journal of the National Medical Association*, vol. 87, no. 3, 1995.

Sattler, Jerome M. *(1988). Assessment of Children.* Third Edition, published by the author. San Diego, CA: Jerome Sattler, Publisher.

Simeon, D.T., and S. Grantham-McGregor. "Nutritional Deficiencies and Children's Behavior and Mental Development." *Nutrition Research Reviews*, vol. 3, 1990, pp. 1-24.

Simeon, D.T., and S. Grantham-McGregor. "Effects of Missing Breakfast on the Cognitive Functions of School Children of Differing Nutritional Status." *American Journal of Clinical Nutrition*, vol. 49, 1989, pp. 646- 653.

U.S. Department of Agriculture and U.S. Department of Health and Human Services. *Dietary Guidelines for Americans*. 4th edition. Washington, DC: U.S. Government Printing Office, December 1995.

Vaisman, N., H. Voet, A. Akivis, and E. Vakil. "Effect of Breakfast Timing on the Cognitive Functions of Elementary School Students." *Archives of Pediatric & Adolescent Medicine*, vol. 150, 1996, pp. 1089-1092.

Wachs, T.D. "Relation of Mild-to-Moderate Malnutrition to Human Development: Correlational Studies." *Journal of Nutrition Supplement*, vol. 125, 1995, pp. 2245S-2254S.

Wahlstrom, K.L., A. Bemis, and J. Schneider. "Minnesota Universal Breakfast Pilot Study: Final Report, Year Three, 1996-97." Center for Applied Research and Educational Improvement, December 1997.

Wyon, D.P., L. Abrahmansson, M. Jartelius, and R.J. Fletcher. "An Experimental Study of the Effects of Energy Intake at Breakfast on the Test Performance of 10-Year-Old Children in School." *International Journal of Food Sciences and Nutrition*, vol. 48, 1997, pp. 5-12.